Virtual patient design in undergraduate education

Dr James Bateman, MBChB, MRCP, MMedEd

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy in Medical Education.

University of Warwick, Warwick Medical School

November 2013

Table of Contents

i.) Acknowledgements

This research project is funded by an Educational Research Fellowship from Arthritis Research UK, Grant number 19330. I am indebted to the charity for the professional and personal support over the past three years.

I would like to thank my supervisor Dr David Davies for supporting the original idea to conduct virtual patient research at Warwick Medical School, for assistance in conceiving and refining the research study from application for funding to supervision, guiding me through my MMedEd, and the initial research component of the PhD.

I would also like to thank Dr Maggie Allen, co-supervisor for help, support, training and guidance from the projects first beginnings in 2009. I would like to acknowledge the research faculty, staff and students from Warwick Medical School who have assisted in this project. Thanks also to the musculoskeletal faculty and students from the two collaborating medical schools, Birmingham and Keele.

Finally to my family: Nina, Alex, Theo and baby Sebastian. Thanks for putting up with me.

ii. Declaration and Signed Statement

‘This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy. It has been composed by myself and has not been submitted in any previous application for any degree.

The work presented (including data generated and data analysis) was carried out by the author. A list of conference papers is included in **Error! Reference source not found.** (p.**Error! Bookmark not defined.**)

Parts of this thesis have been published by the author. They are included in the **Error! Reference source not found.** (p.**Error! Bookmark not defined.**).

Signed statement

I am aware of the University regulations governing plagiarism and I declare that this document is all my own work except where I have stated otherwise.

Signed ……………………….

Full Name ……………………….

Date ………/……/……….

iii.) Thesis Abstract

Title
Virtual patient design for undergraduate medical student education. A prospective, multi-centre research project using qualitative and quantitative methods.

Background
Virtual patients (VPs) are computerised online representations of realistic clinical cases. Recent technology and software advances position VPs as a standardised, accessible, collaborative teaching tool. We do not know how they should be designed. My research question is: how do different VP design principles influence student experiences when completing VPs? The aim of this study is to provide qualitative and quantitative research evidence to support VP design and development.

Methods
This research project uses qualitative and quantitative methods to evaluate how VP design influences medical student learning, based on groups of students from three UK medical schools (Warwick, Birmingham, Keele). The initial qualitative research component is a grounded theory (GT) focus group study evaluating VP design properties. The literature review and qualitative research identified the two most important VP properties to research were: (1) branching within the cases; and (2) structured clinical reasoning instruction (SR) intended to promote good clinical decision making in the VPs. The quantitative research component is a multi-centre randomised experimental 2x2 factorial study of undergraduate students at three UK medical schools, conducted to a published protocol. I investigate two most important independent VP design variables: (1) branching, present or absent; (2) SR, present or absent. Outcomes including: (a) VP scores; (b) VP student evaluations; (c) metrics collected from the VP environment; (d) student self-reported case preferences and (e) summative assessment results. The study has institution ethics approval.

Results
In the qualitative study of six focus groups (n=46), I produced a model describing how VP design influences learning. In the quantitative research, 572 students completed 1773 VPs, and 1223 evaluations, with 296 (50.1%) students completing all four VPs (1184). Key findings were: student expressed preferred SR when present (70.5% of student, $P<0.001$); there were no significant differences in adjusted global VP scores or evaluation scores (all $p>0.3$ for the independent variables); institution factors played an important role with higher scores at one centre ($p<0.001$); and there were significant improvements in Bayesian reasoning with SR present (7% improvement, $p<0.001$).

Discussion
This original research is the first GT study into VPs. The quantitative component is the largest study to date in the literature exploring VP design variables. It provides practical lessons for authors and institutions for design and delivery of VPs. All VPs used are available as open education resources.

iv.) List of abbreviations

AAMC	American Association of Medical Colleges
ANOVA	Analysis of Variance
ANCOVA	Analysis of co-variance
BEME	Best Evidence in Medical Education
CAQDAS	Computer assisted qualitative data analysis
CCTST	California Critical Skills Thinking Test
DTI	diagnostic thinking inventory
EMQ	extended matching question
EViP	European virtual patients project
GT	Grounded theory
IPE	Intermediate Professional Examination (Warwick Medical School)
IQR	Inter-quartile range
JISC	Joint Information Systems Committee (now known as Jisc®)
KFP	key feature problem
KMS	Keele Medical School
MCQ	multiple-choice question
MSK	Musculoskeletal
MVP	MedBiquitous Virtual Patient Standard
OER	Open Education Resources
OSCE	Objective Structured Clinical Examination
QQ	Quantile-quantile
SGUL	St Georges, University of London
SNAPPS	Acronym describing structured reasoning: (summarise case, narrow differential, analyse differential, probe preceptor, plan, self study).
SR	Structured Reasoning
UBMS	University of Birmingham Medical School
UHCW	University Hospitals Coventry and Warwickshire
VLE	Virtual Learning Environment
VP	Virtual Patient
VP Sim	Virtual patient Simulation, University of Pittsburgh
Web SP	Web Simulated Patients, Karolinska Institute
WMS	Warwick Medical School
XML	Extensible mark-up Language
ZPD	Zone of Proximal Development

v. List of figures

v. List of Tables

vi. Appendices

Section 1. Literature Review

In this section I critically appraise the literature on the development of Virtual patients (VPs) in medical education over the past four decades. This section begins with an introduction to the VP, associated software authoring platforms, and recent international technical standards that potentially redefine the properties of a VP. I then outline the evidence supporting their use and their role in education, in the context of developments in e-learning, web-based learning, and open education resources.

1.1. Introduction to 'virtual patients'

No single definition of a 'virtual patient' (VP) is accepted in the literature, with authors and institutions describing them in sometimes subtly different ways. I present three definitions here. The Association of American Medical Colleges (2007) define VPs as:

> *"a specific type of computer programme that simulates real-life clinical scenarios; learners emulate the roles of health care providers to obtain a history, conduct a physical exam, and make diagnostic and therapeutic decisions."*

This definition has been used in two literature reviews on computerised cases and virtual patients respectively (Cook et al., 2010, Cook and Triola, 2009). Ellaway et al., (2008a) define them by a consensus statement as:

"interactive computer simulation of clinical scenarios for the purpose of medical training and assessment"

The MedBiquitous VP Standard (MedBiquitous Virtual Patient Working Group, 2010) defines a VP as:

"an interactive computer simulation of real-life clinical scenarios for the purpose of medical training, education, or assessment. Users may be learners, teachers, or examiners."

These definitions provide the scope, role and development of VPs, but do not precisely define design principles or the technical or software specifications required when authoring, editing or administering a VP case. All of these properties potentially influence the VP use, effectiveness and potential for collaborative use.

1.1.1. The Journey of Virtual Patients to present day

Virtual patients have changed in the past decade. There has been a drive towards entirely web based systems, collaborative licencing of intellectual property, and technical standards for VP authoring and sharing. These three developments have resulted in a paradigm shift, with a fundamental impact on future research strategies in the VP field (Bateman and Davies, 2011).

1.1.2. Developments in authoring platforms for VP cases

By the middle of the last decade a diverse range of VP authoring software platforms had appeared in the literature. This was against the backdrop of spiralling development costs of frequently over $100 000 per VP (Huang et al., 2007) with the use of bespoke single centre/ speciality IT projects to design computer assisted learning cases (Friedman et al., 1991). The newer platforms include: 'CAMPUS', developed by the University of Heidelberg, Germany (Ruderich et al., 2004); 'Labyrinth' from the University of Edinburgh, UK (Begg, 2010); 'Web-SP' from the Karolinska Institute, Sweden (Zary et al., 2006); the 'Tufts University's VP Player' from Boston, USA (Lee et al., 2008); 'VP-Sim' and 'DecisionSim' from the University of Pittsburgh, USA (McGee et al., 1998); 'CASUS' from the University of Munich, Germany (Fischer et al., 1996, Franke et al., 2002), the 'New York University VP player', USA (Triola et al., 2006). Other bespoke VP systems exist that have been created independently of MedBiquitous, such as Harvard Medical School VP programme (Shapiro Institute for Education and Research and Beth Israel Deaconess Medical Centre, 2004). These software developments began to move the practical process of VP authoring from information technology professionals to healthcare practitioners and educators.

The range of authoring options and a lack of standardisation between systems presented a problem for VP research and development. For example some software developed had particular design properties such as the ability to create branching cases, the ability for users to take different pathways through a case depending on the decisions made, adopted by 'Labyrinth', VP-Sim, and DecisionSim (MedBiquitous, 2011). In response to this, in 2005 the organisation MedBiquitous began a process to define a VP interoperability standard for web based VP authoring. MedBiquitous (www.medbiq.org) is a not for profit organisation accredited by the American National Standards Institute (ANSI), set up by leading American Medical Societies.

MedBiquitous is supported by academic institutions and industry, and has a strong partnership with the information technology company International Business Machines Corporation (IBM). The MedBiquitous Working Party to define VP standards first met in 2005, initially including representatives from the University of Pittsburgh, the Karolinska Institute, Tufts University, and New York University. They began to develop a technical specification to encode VP data on the web (Triola et al., 2007). For software to be entirely 'web based' means that case information is stored, accessed and edited online (Zary et al., 2006). The clear benefits of web based systems and standards when developing VPs into a sharable collaborative teaching resource has been highlighted (Ellaway and Masters., 2008b). The consortium defined and standardised the way VP software should 'save' VP information written by an author. They used an existing established, widely used web standard called XML (extensible mark-up language), developed and updated by the W3C (World Wide Web Consortium, 2008). This allows information from a document or programme to be categorised and stored in a systematic way.

1.1.3. The MedBiquitous Standard now defines VPs

The MedBiquitous Virtual Patient Group (2010) set a technical standard for cases describing the core components of a VP (the MVP standard). This standardisation allows sharing of cases between software platforms and institutions. There are four areas into which information is encoded. The four VP areas are: (1) VP case data; (2) rules on revealing data to learners; (3) media resources available; (4) the ways in which a learner can interact with a case, such as question styles. Cases are designed as collections of individual nodes, which can be any combinations of media, such as text, pictures, audio or video, along with tasks. Tasks may be questions, history taking, or clinical decisions. A case with five nodes joined together in a linear fashion is presented to the learner as five steps. Activities within any node may be a set number of questioning styles, or a selection of different decisions. There is the potential to link pathways in a linear or branching style, from simple to complex designs. The MedBiquitous stakeholders and institutions developing software platforms adopted the initial XML prototype. The technical standard was adopted by collaborative research projects such as a European Union funded Collaboration named the 'the European VP Project' or EViP (Smothers et al., 2008). In 2010 MedBiquitous published the MVP XML standard, already supported by many of the web based software development tools (MedBiquitous Virtual Patient Working Group, 2010). I consider the use of the MedBiquitous VP standards as analogous to the standards enjoyed by the widely used PDF file format (portable document format, Adobe Systems). When accessing the content of a PDF, specific functionality of the storage, retrieval, processing, formatting and user tools will differ depending on the hardware and software used to read the email. Nevertheless, the core content and functionality is preserved. A virtual patient XML file is similar. It can be therefore used and played on different virtual patient software platforms, and the core case content (text, media, branching, questioning) will be constant. I have described the basic building blocks of a VP as nodes. I present a schematic of the node types in Figure 1, describing how nodes can be linked to form a VP.

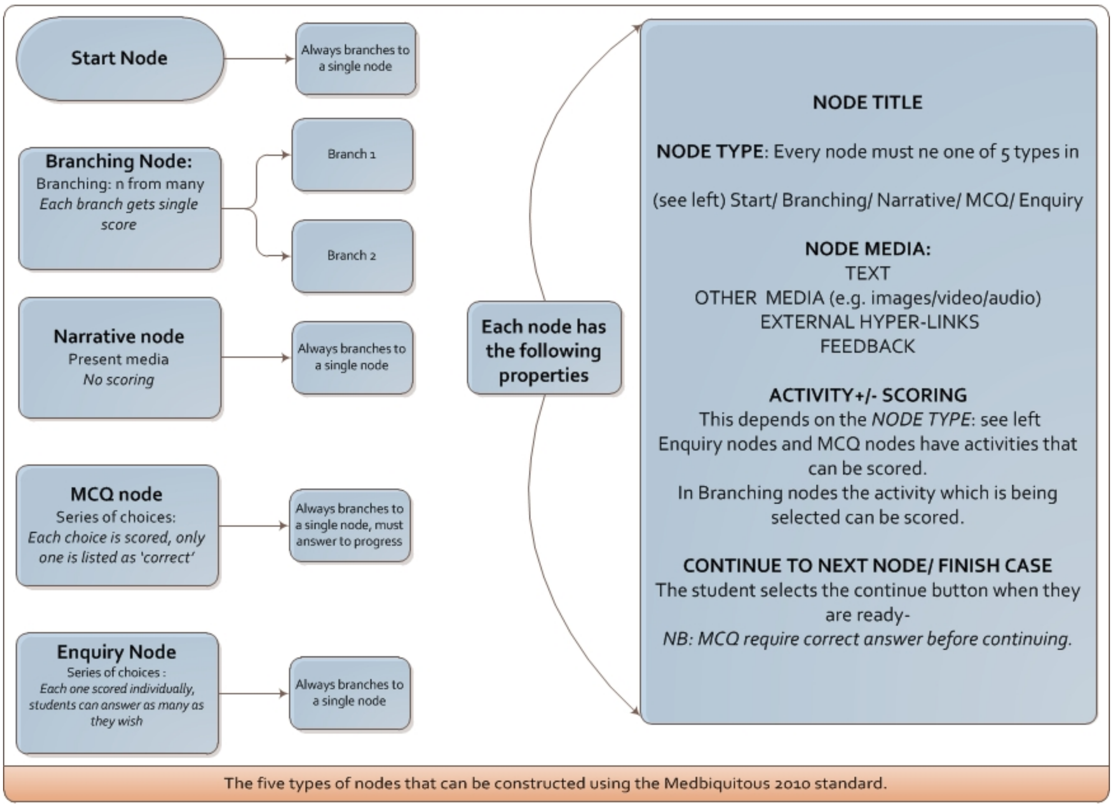


Figure 1 Typologies of nodes, based on the MedBiquitous standard

1.1.4. The Evidence supporting the use of Virtual Patients

The use of VPs is supported by research evidence and educational theory. Virtual patients are a form of web based learning. Web based learning is increasingly used (Issenberg et al., 2005), and was found to be effective in a range of outcome measures in a 2008 meta-analysis (Cook et al.). This included 201 studies, including some research using electronic teaching cases that fulfilled a broad definition of a VP. More recent VP research has provided quantitative and qualitative data outlining the uptake and effectiveness of VPs as a form of web based learning across a range of outcome measures. VPs were already being used in over one-third of US Medical Schools towards the end of the last decade (Huang et al., 2007). A systematic literature review of VPs (Cook and Triola, 2009) cited evidence from a number of domains and health care professions in terms of knowledge, satisfaction and skills gained. Cook and Triola proposed that VPs are ideally placed to teach clinical reasoning skills, citing distinct advantages over both traditional teaching formats and new technologies such as simulation. This was followed by a systematic review and meta-analysis of 54 studies investigating the effectiveness of VPs (Cook et al, 2010), including research published up to February 2009. The research was conducted across a range of undergraduate and postgraduate healthcare professions including medicine, dentistry, nursing, and physiotherapy. VP topics were diverse and included over 30 clinical subject areas from abdominal pain assessment, to communication skills. This research found improvements in knowledge (11 studies), skill acquisition (8 studies), and clinical reasoning (8 studies). It also demonstrated positive student satisfaction scores with VPs (8 studies). It found a large effect size for improvements in knowledge gained. Important qualitative research has subsequently been published. This supports the positive student experience and longer-term knowledge gain from VPs (Botezatu et al., 2010a; Huwendiek et al., 2009b). Although somewhat limited, there is also evidence to support the correlation of VP performance with written examinations (Botezatu et al., 2010c), and support for use for VPs as assessment tools (Round et al., 2009). More recent quantitative research has focussed on knowledge gain in specific subject areas such as psychiatry (Lin et al., 2012) and neurology (Johnson et al., 2013).

Educational theory supports the role of VPs as a form of case based learning. They have long been recognised as being able to facilitate the transitions between basic science education, initial clinical experience and real clinical practice (Voelker, 2003).

They have a pedagogic role simulating experiential learning (Kolb, 1984) and facilitating repetitive deliberate practice (Ericcson, 2004). Kim et al. (2006) synthesised a series of desirable components for case based learning in a in a multi-disciplinary review. They describe a conceptual framework for case delivery which include including case content, structure, and processes that are consistent with the Medbiquitous (2010) VP technical standards used for case authoring. From a pedagogic perspective, VPs are technically well placed to deliver the components outlines by Kim and colleagues. In particular the need for flexibility in case structure and the gradual revealing of information proposed by Kim lend themselves to VP technologies. Following from the work of Kim at al., Cool and Triola (2009) have proposed that VPs have clear strengths over traditional learning technologies in some areas such as clinical reasoning, but significant weaknesses in others such as history taking and clinical examination. Therefore it seems the literature supports VPs as delivering clinical reasoning skills training, promoting hypothetico-deductive reasoning skills (Norman, 2005): this is discussed later in '*1.5* VPs and Clinical Reasoning*, p.52'*. Unfortunately despite presenting evidence supporting the use, uptake, role and benefit of VPs, none of the literature presented here allows us to draw firm evidence based conclusions about how VPs should best be designed (Cook et al., 2010).

1.1.5. Open Education Resources (OER)

The MedBiquitous standard has the potential to promote sharing and collaboration between institutions (Campbell et al., 2009). The United Nations Educational, Social and Cultural Organization (UNESCO) define Open educational resources (OER) as:

“Open educational resources (OER) are materials used to support education that may be freely accessed, reused, modified, and shared.” (UNESCO, 2011)

Collaboration within institutions brings challenges surrounding intellectual property copyright. Addressing these is vital for any research project intending to produce open access resources (Mayor, 2013). Several advances in the licencing of academic work have addressed these challenges. The most widely used in the medical sphere is the Creative Commons Project (US National Library of Medicine and National Institute of Health, 2012) originally devised by Lawrence Lessig of Stanford Law School (Gould, 2004), and available on the Internet from http://creativecommons.org. This presents a legal framework for sharing intellectual property as stipulated by the original author using four core provisos: attribution; commercial/ non-commercial use; licence to modify the original work; and stipulation to share original and derivative works under the licence specified by the original author. This means authors of research cases can share intellectual property and remain the recognised author. These standards have been successfully applied to VPs (Smothers, 2008).

1.1.6. Web Based Technology and Virtual Patients

Web based learning is a form of e-learning. There are many different interpretations of what web based learning is.

Perhaps its broadest definition is:

“the use of the Internet for education” (Ellaway and Masters, 2008b)

VP cases today are authored, stored, played and evaluated exclusively using the web, which is increasingly available. This development represents a paradigm shift from previous case based learning described in the early authoring platforms.

Web based authoring means that cases are stored in an online repository. Importantly this means authors and learners can access and play cases using computers connected to the Internet requiring no bespoke downloaded software. Cases can be downloaded in the form of a single case file, and distributed electronically. The online environment also allows the real time administration of cases. Inter-operability web standards extend to cross platform computer operating systems. Access to the web in the UK is close to universal. In 2009 over 85% of medical students had home broadband Internet access and a computer in the UK, and all had 24-hour access via their studying institution (Khan et al., 2009).

1.1.7. The evolution of virtual patients: shifting real world definitions

The technology shifts and technical standards discussed above represent an important evolution in VP development from three different perspectives. Firstly the VP technical standard impacts on the 'real world' definition of what a VP is (Bateman and Davies, 2011). Secondly, technical standards mean cases can be created by different software packages in different institutions, however all cases created to the standard are interoperable, facilitating sharing and collaboration. Finally, new technology places the clinician as the author, editor, publisher, and sharer of the case rather than an information technology professional. My schematic representation of these three elements is shown in Figure 2.

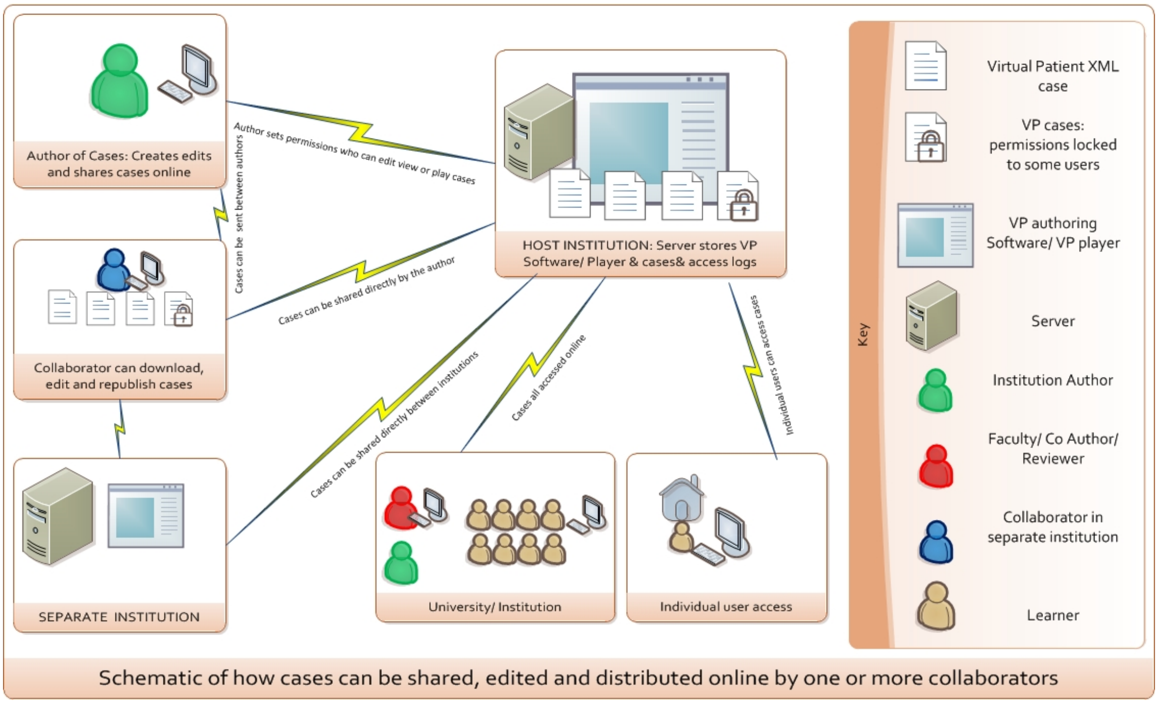


Figure 2 Schematic representation of how VPs are shared between authors, collaborator and institutions

I argue this means that the VP today is fundamentally different in its origins from the commonly cited early descriptions of the first 'VPs' in the early late 1960's (Bitzer, 1966, cited by Cook et al., 2010) and 1970's (Harless et al., 1971, cited by Cook and Triola, 2009). Certainly this early computer assisted learning research represented the leading edge of development in the field at the time of publication, and contains valuable research information. However I will outline in the coming sections how the this early research bears little resemblance to what virtual patients *actually are* today, how they are used, authored or designed.

1.2. Literature review

In this sub-section I systematically and critically assess contemporary research evidence for VPs. I begin by describing the landscape of the research, then a focussed literature review from 2000 to the commencement of the study. I have identified eight themes: historical perspectives, authoring, knowledge gain, students, VP adoption, feedback in VPs, and the wider e-learning literature. At the end of this section I discuss the rationale for the review strategy.

Two important literature reviews have coincided with the commencement of this research. The initial literature review in 2009 (Cook and Triola) was used to support the grant application for this research. The second, “Computerized virtual patients in health professions education: a systematic review and meta-analysis”, the largest and most comprehensive review on computerised teaching cases was published by Professor David Cook, a professor of medicine and medical education at the Johns Hopkins Institute, at the time I began my research (Cook et. al., 2010).

As a result of the comprehensive literature review published by Cook and colleagues at the beginning of my thesis, a focussed review of recent literature on VPs was conducted. The objective of this focussed literature review was to examine the VPs design properties used and evaluated in contemporary VP research. I followed an established approach for healthcare (Centre for Reviews and Dissemination, 2009). My search strategy was as follows. Three databases were searched: OVID Medline, PubMed, and EMBASE, described in Figure 3 (p.32). The PICO (population, intervention, control, outcome) approach was used to define the research question

(Schardt et al., 2007). My population were medical and dental trainees at both undergraduate and postgraduate levels. I defined a ‘contemporary VP’ for this review as a computerised learning experience described as a ‘virtual patient’. This had to either (1) fulfil the AAMC definition (see Section 1.1) or (2) potentially capable of compatibility fulfilling the MedBiquitous Virtual Patient Working Group (2010) XML standard.

I excluded studies that evaluated cases that cannot be adapted to be conceivably shared between institutions under the MedBiquitous standard. This includes bespoke technology such as 3-dimensional learning environments, haptic simulation, 3D worlds, voice recognition, and natural language processing. I also excluded: studies published before 1st January 2000, studies published in a language other than English; conference proceedings without full text online abstracts; oral presentations; studies involving other allied professions (vetinarians, paramedics, and the like); and any study where the full text was not available (see Figure 3). I did not exclude other journal articles such as letters, commentaries, or narratives. The summary of the papers identified for review is shown in Figure 3.

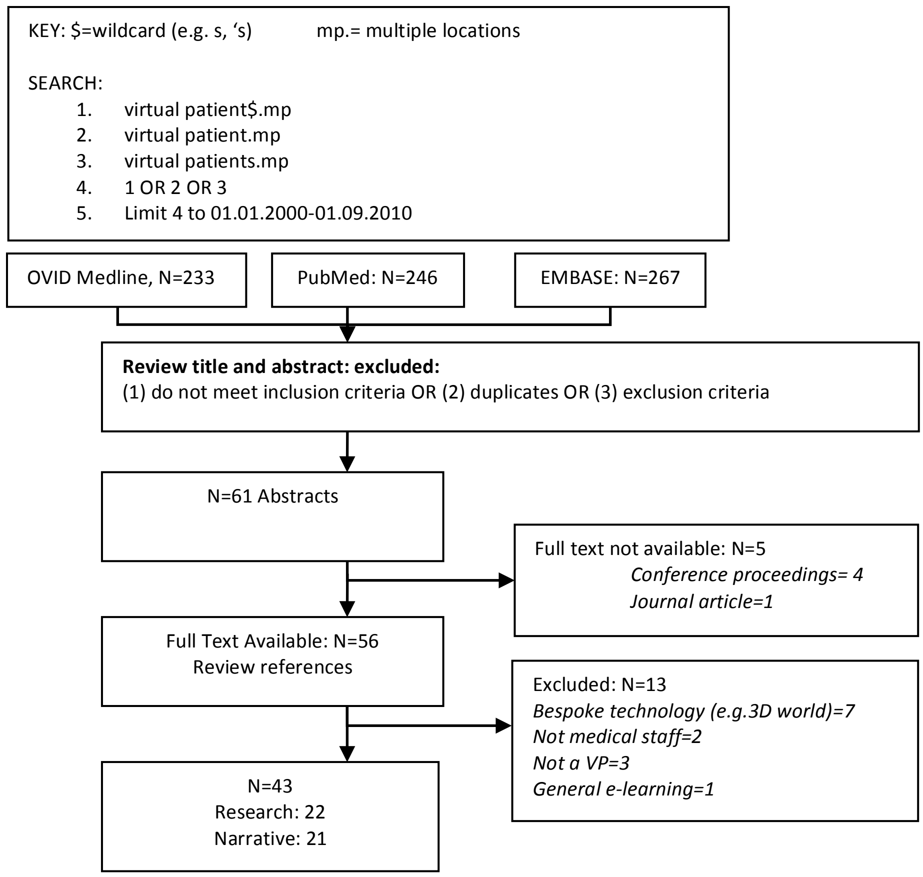


Figure 3 Search strategy for virtual patient literature

From the 61 potentially appropriate abstracts, with full text being available for 56-in total 43 papers met the inclusion criteria. I used a nine point checklist for assisting the critical appraisal of the papers (Morrison et al., 1999). I have summarised the reviewed abstracts in Table 1. I used a simple descriptive classification system to classify them in terms of their primary focus into ten categories based on the feature variations of VPs and research described by Cook in his first 2009 review. The categories of the research papers I evaluated included VP design; VP feedback; curricular integration; knowledge transfer and retention, assessment; role in different specialties; VP typologies; historical perspectives; VP software tools; and cost. For experimental studies, I have described their impact using an established hierarchy (Kirkpatrick, 1967). Kirkpatrick level 1 measures reaction, level 2 measures learning, level 3 measures behaviour change, level 4 measures outcomes. Level 2 is divided into 2a (modification of attitudes and perceptions) and 2b (modification of knowledge and skills).

Table 1 The 61 research publications identified, highlighting the 18 exclusions, the complete references are shown in appendix 2.

Suitable Abstracts [N=63]	Included: full text review [N=43]	Type: R= Research N= Narrative	Summary	Participant & Clinical topic notes	Primary area
Botezatu, 2010	Y	R	Experimental design randomised two group comparative study exploring knowledge retention in VPs compared to traditional methods (N=52), favoured VP. Kirkpatrick 2b	MS, haematology, cardiology)	Knowledge retention
Huwendiek, 2010	Y	R	Descriptive study. Description of several evaluation tools for VPs, open access	MS (general)	Evaluation
Tan, 2010	Y	R	Descriptive case Study from three US schools, including sample cases. Evaluation tool presented (N=37). Kirkpatrick 2a	MS, general/ elderly care	Evaluation, feedback, VP design
Lehman, 2010	Y	R	Non-comparative, non-randomised single group post intervention study (N=30). Studied curricular integration in a blended learning using VPs alongside clinical skills, Kirkpatrick 2a	MS, clinical skills	Curricular integration.
Tworek, J., 2010	Y	N	Narrative study: expert opinion, observation, narrative: cites evidence, descriptive, conceptual	MS, ED	VP design, knowledge transfer, Feedback
Nimoyama, S. 2009	N		Non-medical staff, evaluated technicians.	N/A	N/A
Horstmann, 2009	N		Descriptive study, N=70. Evaluates virtual hospital environment (to teach urology). Post-test student satisfaction only	MS	
Graham, 2009	N		Not a VP study		
Conradi, 2009	Y	N	Narrative Expert opinion and review.	MS	Assessment
Poulton, 2009	Y	N	Narrative expert opinion and review	MS	Use of VPs
Huwebdiek, 2009	Y	R	Focus group study of design principles for VPs (N=27). Thematic analysis. Kirkpatrick 2a	MS, paediatrics	VP design, feedback
Huwendiek, 2009	Y	N	Expert opinion and review	N/A	Typology
Gesundheit, 2009	Y	R	Single group experimental study evaluating student satisfaction. (N=27) Kirkpatrick 2a	MS	Satisfaction
Fors, 2009	Y	N	Narrative, Expert opinion.	MS	Integration
Dewhurst, 2009	Y	N	Narrative, Expert opinion	MS	Integration
Conradi, 2009	N		Paramedics, virtual 3d world		
Campbell, 2009	Y	N	Narrative, expert opinion	N/A	Integration
Ellaway, 2009	Y	N	Narrative, Expert opinion	N/A	Historical Perspectives
Kim, 2009	N		General e-learning in Korea		
Zary, 2009	Y	N	Descriptive study, expert opinion, cites evidence	N/A	Interoperability
Kononowicz, 2009	Y	N	Descriptive study, expert opinion	N/A	Interoperability
Bardella, 2009	Y	N	Narrative, expert opinion	MS	Assessment
Kenny, 2009	N		Bespoke VP: Uses 3D world		
Berman, 2009	Y	R	Multi-centre post test experimental study focussing on curricular integration (N=545). Kirkpatrick 2a	MS	Curricular Integration
Posel, 2009	Y	N	Expert opinion, cites evidence.	N/A	VP design, feedback
Huwendiek 2009	Y	R	Qualitative research study, focus groups, thematic analysis (N=27) Kirkpatrick 2a	N/A	VP design
Nirenmerh, S.	N		Bespoke VP (natural language processing)	Zary,	
Cook, 2009	Y	N	Critical literature review and proposed steps.	MS/ PS	Literature review
Zary, 2009	Y	R	Descriptive single group study (N=99) evaluating 23 VPs with or without feedback. Kirkpatrick 2a	DS	Feedback
McConnell, 2008	Y	N	Narrative of available VP design software platforms, and historical research in VPs	MS	Authoring platforms
Deladisma, 2009	N		3D virtual world, main focus is manikin for breast examination.	N/A	N/A
Smothers, 2008	Y	N	Descriptive study, expert opinion on sharing VP cases.		sharing, Interoperability
Sumner, 2008	N		Conference poster, unavailable	N/A	N/A
Sumner, 2008	N		Conference poster, unavailable	N/A	N/A
Critchley, 2008	Y	R	Single group post-test study, Student satisfaction. Numbers of students not listed by authors. Kirkpatrick 2a	MS, anaesthesia	Satisfaction
Hooper, 2008	Y	R	Factorial study design employing mixed qualitative and quantitative methods. N=404 GPs (USA). Note use of bespoke technology (microphone, GPs reading pre-scripted questions to computer). Complex methodology (described as '2x2x2x2x2x2 study design') Kirkpatrick 2a	GP, depression assessment	Student population, VP use

Suitable Abstracts [N=63]	Included: full text review [N=43]	Type: R= Research N= Narrative	Summary	Participant & Clinical topic notes	Primary area
Triola, 2007	Y	N	Expert opinion	N/A	Interoperability
Dev, 2007	N		Conference poster, unavailable	N/A	N/A
Vukanovic-Criley, 2008	Y	R	2 group experimental study, non-randomised study. Note: does use some bespoke VP technology (speakers to work with stethoscopes). Favourable outcomes for intervention (N=24) over standard teaching (n=58). Kirkpatrick 2b	MS, Cardiology	Role, VP use
Courteille, O	Y	R	Pilot single group study of VPs as assessment (OSCE) tools. N=118	MS, OSCE station	Assessment
Ellaway, 2008	Y	N	Expert opinion. Narrative on sharing VP cases	MS/ other	Interoperability
Bittner, 2008	N		Bespoke 3D technology		
Waldmann, 2008	Y	R	Experimental single group study describing assessment in final year medical students (N=146). Scoring refers to PhD thesis published in German.	MS, general practice	Assessment
Orton, 2008	Y	R	Descriptive study of 287 student evaluations from 16 institutions from the University of Iowa. The student evaluation is not presented. Kirkpatrick 2a	MS, postgraduate. Geriatrics	Student satisfaction
Ghosh, 2007	N		Conference proceedings not available online		
Orr, 2007	N		Excluded: Pharmacy students, N=123	Pharmacy students.	Student satisfaction. Integration.
Deladisma, 2007	N		Bespoke 3D VP case		
Huang, 2007	Y	R	Descriptive study citing evidence of VP uptake and costs at US institutions	MS	Integration; cost
Vash, 2007	Y	R	Randomised (method unclear) post-test study of Iranian medical students (N=48) completing VPs. Methods not validated. Post-test scores suggest benefit from VP. Kirkpatrick 2b	MS, general surgery	Skills training, Assessment
Triola, 2006.	Y	R	Randomised Experimental study (two groups) comparing VPs to live clinical actors in practicing family medicine doctors in the US (N=55) . No inferiority seen in VPs compared to actors. Kirkpatrick 2b	GPs, Mental health	Student satisfaction, VP Use
Zary, 2006	Y	R	Descriptive study and pilot evaluation of VPs for dental, medicine and pharmacy students (N=220 returned evaluations). Kirkpatrick 2a	DS, MS, PS	Student satisfaction
Dickerson, 2006	N		Bespoke technologies (synthesised speech)		
Walsh, 2005	Y	N	Narrative from BMJ learning editor on bespoke e-learning module development.	Postgraduate trainees	Student satisfaction; interoperability
Simo	Y	N	Expert opinion describing an approach for teaching cardiology emergencies	MS, cardiology	VP design
Ruderich, 2004.	Y	N	Description of a VP player, CAMPUS	MS	Interoperability, VP design
Marco, 2004	N		Paper unavailable, PubMed abstract=36 words.		
Voelker, 2003	Y	N	Narrative drawing from VP use in a number of institutions including Hull Medical School, Minnesota University and international institutions		VP use.
Malamateniou, 2003	N		Study of electronic patient records		
Bearman, 2003	Y	R	Qualitative study, interviews (N=12). The author extracted 23 themes from students who used the virtual patients. Kirkpatrick 2a	MS, cardiology	VP use; skills training; VP design
Bearman, 2001	Y	R	Mixed methods study evaluating ‘problem solving’ and ‘narrative’ VPs in one centre (N=267). Small significant differences (0.3 on 5 point Likert scale). Uses 10 item simulation assessment tool. Focussed on videos of cases. Kirkpatrick 2a	MS, Cardiology	VP Design, evaluation.
Bearman, 2001	Y	R	Description of the same studies, students, case, and methods as the other study published in 2001. Kirkpatrick 2a	MS, Cardiology	VP Design, evaluation

Abbreviations

Y,yes; N, no; R, research; N, narrative; MS, medial student; GP, general practitioner; ED, emergency department,

I have used this review to present the literature on eight contemporary research areas: (1) evolving typologies; (2) VP design research; (3) VP feedback research; (4) different trainee levels; (5) curricular integration and sharing; (6) VPs as assessment tools; (7) knowledge retention; and (8) development costs. I conclude this subsection with a discussion of research into bespoke computer cases that I have not included in this review, alongside a summary of the evidence. Some of the papers form part of the historical introduction to VPs already presented (interoperability, technical standards, role).

1.2.1. VP typologies: evolving in the last decade

Dr Margaret Bearman is a Health Informatics trained Senior Lecturer from Monash University, Australia who led early VP typology work. She should not be confused with Dr Norm Berman, a paediatric cardiologist and director of InTIME (Institute of Innovate Medical Technology from Dartmouth University, USA, see Section 1.2.5, p.42). Bearman et al. (2001) proposed that two different VP design principles exist: a 'problem solving design'; and a 'narrative design'. A 'problem-solving' design describes clinical reasoning following an established rule based model. A narrative design is described as a bespoke case where the "*narrative has to be individually crafted rather than relying on a template*" (p.1004). Bearman's research found no differences between case designs using a questionnaire that had not been validated, however Bearman does comment on some qualitative findings.

> *"Analysis of student responses suggested that the narrative simulation was more encouraging of a reflective process …the problem-solving simulation*

was perceived as providing benefits with respect to use of appropriate language." P.1008

In further qualitative work by Bearman (2003), using a video based model of history taking, the authors propose that problem solving designs are easier and less costly, despite being associated with more video clips (234 for problem solving, 154 for narrative case). Although the descriptions are useful, no exemplar or schematic is presented to explain the organisation of the questions and answers.

The differences between a problem solving and narrative design of VPs (Bearman et al., 2001) were proposed as being clearly delineated as late as 2006 (Zary et al.). It is accepted in the literature that these are now historical, superseded by clear design typologies described by Huwendiek et al. (2009a), see Figure 4, Figure 5. In other VP research in the past 14 years I have found no example of VP research (PubMed and EMBASE) that either (1) clearly represents a schematic of the design of a research case or (2) allows the user to access the VP used. At times interpretation is difficult. Bearman and colleagues (2001) describe the narrative case as a "linear scripted storyline", but also allude to different pathways through the cases: "There was a clear story line, no matter which path the student took" (Bearman, 2003). In this research Bearman identifies 23 themes that describe how students felt when completing a VP (in 1998-99). These represent abstract concepts, such as 'Self-identification as a medical student'. No clear evidence supporting VP design typologies can be taken from Bearman's typology research, or thematic analysis. Bearman does suggest some feedback strategies.

1.2.2. VP Design Research

The most detailed study into VP design in the literature is from Huwendiek et al. (2009b), who conducted a qualitative experimental study to identify effective features of VP design developed using the CAMPUS player (Ruderich et al., 2004). They made 10 recommendations for VP authoring. Although many authors make general comments on VP use and design (Tworek et al., 2010, Bearman, 2003, Voelker, 2003) I discuss Huwendiek's paper in detail to present some of the challenges of researching design features.

The research was conducted on 27 students selected from a cohort who had completed cases with different design properties. The cases were presented in two formats, 'rich content' and 'standard content' in two different modes (text and other media in a graphical user interface, the other text only). The ten design aspects suggested are: (1) relevance; (2) appropriate difficulty; (3) use of feedback; (4) interactivity; (5) appropriate use of media; (6) focus on learning points; (7) recapitulation of key learning points; (8) authenticity of the interface; (9) authenticity of the case (10) questions and explanations to enhance clinical reasoning. I would argue these design aspects are largely generic statements with perhaps the exception of points seven and ten. The data analysis is not explicitly stated: it appears to be a thematic analysis. The authors describe promoting clinical reasoning instruction as helpful, citing three strategies: encouraging hypothesis generation; case abstraction; identification of defining features of the case. The authors do not present examples of how this is done.

In common with the other research described, there are no VP cases nor case schematics (presumed linear structure), media use is not presented (the number of clinical images, radiographs, videos), question structure is not described (e.g. multiple choice, extended matching), quantitative data on patterns of use and case length is not presented, the VPs are not available to the reader. This makes findings supporting feedback difficult to interpret. For example it is not clear if a student is given the option to continue with a decision that is not the correct answer. The process of selecting the number and content of the recommendations is not clear, the authors selected ten practice points. Despite the inevitable limitations of research in this emerging field, this study provided the first qualitative evidence to help to understand how design principles influence the effectiveness of VPs.

Despite producing interesting findings there are a number of factors that limit the impact of this study, analogous to the earlier research by Bearman.
Comparisons in the work include: text interfaces with graphical user interfaces; videos with an unspecified alternative; relevant images with no images. These are perhaps an out-dated approaches when adopting web based teaching practices (Mayer, 2010), going against the principles of best practice against how we should best deploy media (Cook et al., 2008). Clarity is important when describing comparisons with a video:- for example it is not clear if the comparison is a screenshots, diagram, or picture. Comparing a text interface with screen titles such as “Diagnosis-Therapy-Loop-1” represents unnecessary cognitive load, going against the ‘redundancy principle’ (Sweller, 2005) where unnecessary information should be removed. We know students appreciate clinical skills videos for procedures like a

lumbar puncture (Taitz et al., 2006). The authors took an interesting approach of highlighting abnormal findings in the presentation. This approach potentially conflicts with clinical reasoning literature where it is important for the student to evaluate a large numbers of pieces of information and weighing up their importance, using either hypothetico-deductive models (Norman, 2005), and contemporary clinical reasoning theory (Croskerry, 2009b). This is discussed further in Section 1.5.1, 'Clinical reasoning and virtual patients.'.

1.2.3. Feedback in VPs- research

Feedback has been cited as the most important feature of medical simulation in a Best Evidence in Medial Education (BEME) review by Issenberg et al. (2005), and suggested to be important by Huwendiek et al. (2009b). Zary et al. (2009), has researched feedback. A number of authors describe feedback as important, and propose best practice and research areas (Tworek et al., 2010, Posel et al., 2009).

Zary and colleagues deliberately studied feedback in VP cases in a single group experimental study, collecting student feedback across six domains and a global score using 24 VP cases, using the Web-SP system. The authors provide convincing evidence to support the use of feedback in VPs when compared to 'no feedback'. Again, sample cases, scoring systems, the format of the feedback (length, content, triggering, immediacy) are only briefly discussed, and the use of feedback is already well established, its benefits are not surprising. Posel et al. (2009) suggest including assessment and feedback from the start of a case in a largely theoretical descriptive article suggesting tips for VP development. Bearman et al. (2003) highlight the different types of feedback possible in a narrative. In a critical review of the VP literature, Cook and Triola (2009) also propose that feedback is an important area for future study, but are critical of the evidence base to support the use of different feedback modalities.

1.2.4. VPs use in different subject areas and stages of medical training

The majority of research in VPs has taken place in a number of undergraduate specialties, however research on VPs in postgraduates also exists, and both are now discussed. Descriptive studies of VPs in undergraduates describe their use in geriatric education at the University of Iowa (Orton and Mulhausen, 2008), and in prescribing for pharmacy students (Orr, 2007), dental students (Zary et al., 2006),and surgical students (Vash et al., 2007). These studies focus on the role of novel new technology over a detailed description of the design of cases. None of the research allowed the reader to access an exemplar. VPs have been evaluated in some postgraduate studies. A study in primary care physicians in the US evaluated VPs against simulated patients in an apparent linear case design using four different cases(Triola et al., 2006). The software, case structure, feedback techniques and specifics of the case design are discussed only briefly, with outcomes including evidence based interview analysis techniques along with self reported metrics. No significant differences between the groups were found.

1.2.5. Integration of VPs into Curricula

A leading research paper on VP integration was published on the integration of paediatric teaching using VPs into six US Schools (Berman et al., 2009). This is the only research example of integrating cases, although best practice is described by a number of papers (Fors et al., 2009, Smothers et al., 2008).

Berman and colleagues evaluated the success of the Computer Assisted Learning in Paediatrics (CLIP, Institute for Innovative Technology In Medical Education (InTIME), 2012). Berman el al. (2009) surveyed 545 students from six paediatric schools using a tool they had validated to judge the success of the VP cases with the purposes of identifying the best integration strategy. They concluded that the cases should be integrated into courses in replacement of rather than in addition to scheduled activities. The researchers judged VP integration by scoring how many of the following resources were 'eliminated' or replaced by VPs: Textbooks; Other computer-assisted instruction; other examinations; other assignments; didactic hours. The researchers found integration of VPs by removing other resources resulted in self-reported improvement in skills and satisfaction with the cases. This built on earlier work promoting integration strategies into clerkships in the US(Berman et al., 2008). This work (CLIPP) represents the most successful integrated strategy for VPs that has been published (Berman et al., 2011), although the success of the integration of other open access cases has not been published to date (European Virtual Patients Project Development Team, 2010). The use of follow up sessions to reinforce learning has also been shown to improve the student satisfaction of students completing VPs in a Swedish Cohort (Edelbring et al., 2012), and supported the role of feedback, but not assessment in the curricular integration.

1.2.6. Evaluations and Assessments

Several 10-12 item Likert checklists for evaluating attitudes (Kirkpatrick 2a) knowledge and skills (Kirkpatrick 2b) exist (Huwendiek and de Leng, 2010, Zary et al., 2006, Bearman et al., 2001), none have been formally validated. Virtual patients have been studied as assessment tools, although only limited data exists, largely in a formative setting. It has been suggested that VPs can play a role in assessment, particularly in high stakes exams (Round et al., 2009), although only pilot data exists in the literature for their use in assessments such as objective structured clinical examinations, OSCEs (Courteille et al., 2008). A study of nursing students found they would be satisfied to have VPs as an assessment tool (Forsberg et al., 2010), a finding echoed by medical students completing VPs (Gesundheit et al., 2009). Validity evidence is limited to small studies. VPs have been investigated as a form of assessment in haematology and cardiology, and electronic VPs compared to paper based representations of virtual patients (Botezatu et al., 2010c). The researchers used the VPs to assess students who had both been completing VPs during a clerkship and those who had traditional teaching methods. The study design involved students being randomised to learning by completing VPs, or traditional teaching. All students then completed an examination using both a VP and a written paper based examination simulating the content of a VP. The descriptions of the VP do describe some of the scoring items, positively and negatively marked for items such as diagnosis and management. The authors do not include the number of questions, the type of questions, the length of the cases, or number of steps. The cases were reported as linear design. Results were reported for end of block assessments as standardised scores out of 6 for both the written paper and VP.

There appear to be significant methodological flaws in the study design, with students allocated to the VP group scoring at times three-fold higher than those in the traditional group in a written assessment (3.7/6 vs. 1.1/6). The authors found that students taught using VPs performed significantly better on assessments using VPs. The authors acknowledge they cannot adequately account for this, but do state that paper based 'virtual patients' were not positively received by students. In summary, assessment using VPs remains a subject for further research.

1.2.7. Knowledge Retention appears to occur with VPs

There is some work supporting longer-term knowledge gain with VPs. Botezatu et al. (2010b) studied VP simulation and performance in written and online assessments to assess knowledge gain, in a small randomised, non-blinded experimental study of 49 medical students in Colombia studying cardiology and haematology using the Web-SP system. Again the study reporting is of apparently linear VP cases with no branching structure. The analysis is supported with a sparse description of VPs, with no description of structure, length, nodes or media used. Assessments of reasoning used a complex but transparent scoring rubric that has not been validated. The suggestion in the research is that students in the VP group have had practice at using a complicated bespoke scoring system, acting as a confounder. Nevertheless the study suggested significantly better scores in the VP group. The evidence for long term knowledge gain with VPs is limited.

1.2.8. Evidence Supporting Falling Development Costs

The cost of an educational resource (VP) is critical to its uptake. Research in 2007 in the United States suggests one third of medical schools are utilising some form of a VP case, however at significant design cost. Most commonly this was between $10 000 - $50 000 per VP, with a significant proportion costing in excess of $100 000 to develop (Huang et al., 2007). The authors suggest prohibitive costs are a barrier between collaboration and sharing between institutions. The cost implications for the newer generation of web-based VPs are certainly lower, when an author can develop a case in a matter of hours using web based software (Smothers et al., 2008).

1.2.9. Exclusions: Bespoke computer cases

Recent technical standards from MedBiquitous (2010) give a precise definition of a VP. I have used these standards as part of the definition. Any number of computer based e-learning activities could be considered to be close to learning from VP case. I have not included in this review research into several areas of computer-based learning, and I present some areas excluded along with examples of research. I have not evaluated research into: natural language processing, where students interact authentically with patients (Hubal et al., 2000); three dimensional worlds where the user controls a character in an immersive three dimensional environment (Kiss et al., 2004); electronic surgical simulators (Foo et al., 2009); or other bespoke computer simulation. Whilst each of these areas represents a potentially important research field, I consider them to be outside the scope of a VP as defined by MedBiquitous

(2010). I do not consider a slideshow accompanied by an audiotape to be a VP (Mullaney et al., 1976), although this research has been included into meta-analysis into computerised case based learning. Whilst undoubtedly relevant at the time, this practical exclusion is intended to reflect the sea change in VP development today.

1.2.10. Discussion of literature not included in the review.

The review methodology did not encompass published PhD theses, or papers only indexed in certain databases e.g. CINHAL, Scopus or PSYCHinfo. The search does not include computerised e-learning cases that have not been described as 'virtual patients'.

There are examples of computerised teaching cases I have not included early VP research, which however this was covered in a systematic review of the literature (Cook et al., 2010). Cook does cite one paper focussing on MSK education (Wilson et al., 2006), which described rheumatology teaching cases delivered using bespoke web based technology, making cases accessible to readers. Wilson and colleagues (2006) found the cases to be evaluated positively. This research was also funded by Arthritis Research UK (formerly the Arthritis Research Campaign) and took place in a collaborating centre for the second phase of this research (Birmingham).

The semantics of defining or arguing precisely what makes a VP in this historical research is replaced here by a more practical and pragmatic discussion on research evidence supporting VP design. Earlier work in case based computer-assisted simulation evaluated whether we should use computer-assisted learning, and understandably did not focus on design typologies. In one leading example, Mullaney et al. (1976), reviewed the response to naturally typed text using computer assisted technology in dental undergraduates in comparison to a slideshow with an accompanying audio-tape. This group found students did have preferences for different instructional typologies, but these did not follow any identifiable pattern. Work in the 1990's did examine VP design, however there were many limitations in research reporting. I present two leading experimental studies as an example from this decade. Friedman et al. (1991) undertook one of the first detailed examinations of a comparison of computer case based instructional design variables. In a rigorous methodological design they evaluated different instructional design variables in a single computer case. Medical undergraduates (N=72) took part in a randomised post-test experimental study evaluating three difference design properties. Design variables used included a concept of forced clinical reasoning adoption, via menu driven case information with accompanying feedback, or by a natural language enquiry. The student's self-reported experiences indicated they significantly preferred a menu driven format with free text, however no other major significant differences were found in a complex written assessment. No information is given on the case structure (presumed linear), or the content of the menu items. It is implied they are simply revealing particular aspects of a clinical case. Friedman does not report the length of time students took to work through a case; the number of steps; presentation format (text, images); the basis for the algorithms used to identify critical aspects of the history (for example eliciting a history of depression). Assessments do not seem to have been included in the case. The conclusion was that instructional design may influence learning, calling for further work. In the second study variations in case based learning designs were evaluated by Lyon at al. (1991). They described students completing 15 cases around a core topic, but found no significant differences in post-test studies between an interactive computer case, and traditional methods. A major focus was the difference between 'online' help and a textbook. Friedman himself (1994) called for a shift towards the focus on computer assisted learning design over similar comparisons to no intervention or alternative teaching, echoed a decade later in the paper "The research we are still not doing- an agenda for the study of computer based learning" (Cook, 2005b). Cook then extended this to a meta-analysis of web-based learning (Cook et al., 2008), finding it effective . Unfortunately the literature from 1990-2000 addresses predominantly '*If*', but not '*how'* VPs should be designed.

1.2.11. Summary of VP Research evidence

In summary the literature contains numerous examples of VPs in different areas, however no direct comparisons of major typologies such as branching and linear case designs. No studies compare an intervention to improve clinical reasoning skills as an independent variable. To quote Cook and Triola (2009):

> *"Potential variations in VP design are practically limitless, yet few studies have rigorously explored design issues"* (Cook and Triola, 2009), *p. 303*

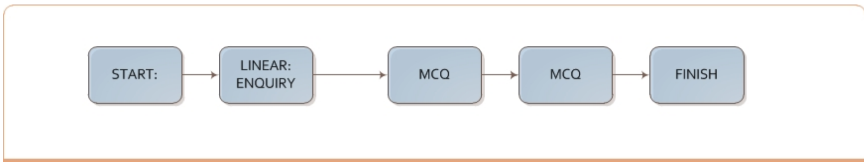
The evidence can be categorised as largely supporting the use of VPs to teach undergraduate and post-graduate students. Cook and Triola (2009) go on to state:

“Unfortunately, very little research has studied these design permutations. As a result, educators have received little guidance in how to develop effective VPs” (Cook and Triola, 2009), p.307

1.3. Open Access VP Case Review:

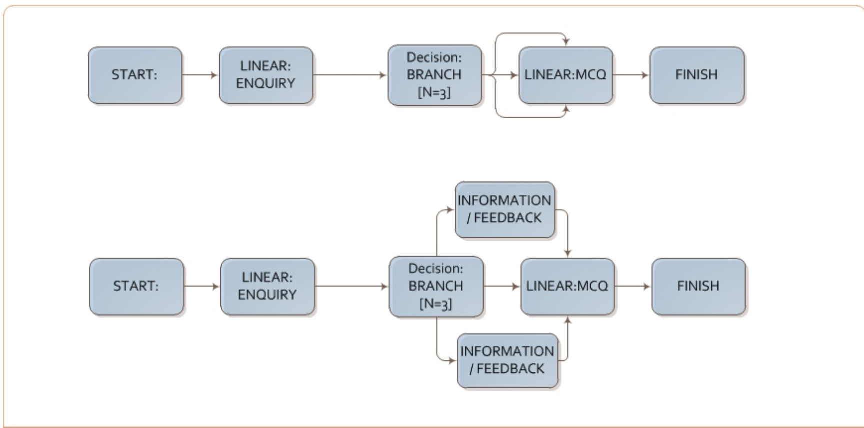
With the evidence I have presented above indicating limited research evidence supporting design, I appraised a number of open access VPs. From the literature review I have identified open access cases from a number of sources. A number of these were no longer available at the time of the review (for example Wilson et al., 2006). I evaluated cases from Harvard Medical School (Shapiro Institute for Education and Research and Beth Israel Deaconess Medical Centre, 2004), St Georges Medical School (St Georges Univeristy E-learning Unit, 2010), the Canadian collaborative open access case project PINE (Northern Ontario School of Medicine, 2010), and the European Virtual Patients Project, or EViP (European Virtual Patients Project Development Team, 2010), to which I have contributed cases (Bateman J. on behalf of the European Virtual Patients Project Team, 2010). EViP cases were authored using a number of the VP case development programmes including DecisionSim, CASUS and CAMPUS software (MedBiquitous, 2011). I have presented the building blocks of VPs in Figure 1, p.24, and the accepted typologies of VPs that can be built from these blocks in Figure 4 (p.50), showing linear and branching cases) and Figure 5 (p. 51), showing the ‘wheel and spoke’ arrangement. The vast majority of open cases use linear navigation (Harvard, St Georges, PINE, all CAMPUS cases) with branching only used in a limited number of EViP cases (they correspond to Figure 4a and Figure 4b), along with the wheel and spoke design (see Figure 5). The design of the cases otherwise conforms to the typologies (Huwendiek, 2009a) shown in Figure 1. This evaluation is limited by the lack of a standardised VP evaluation tool.

a)



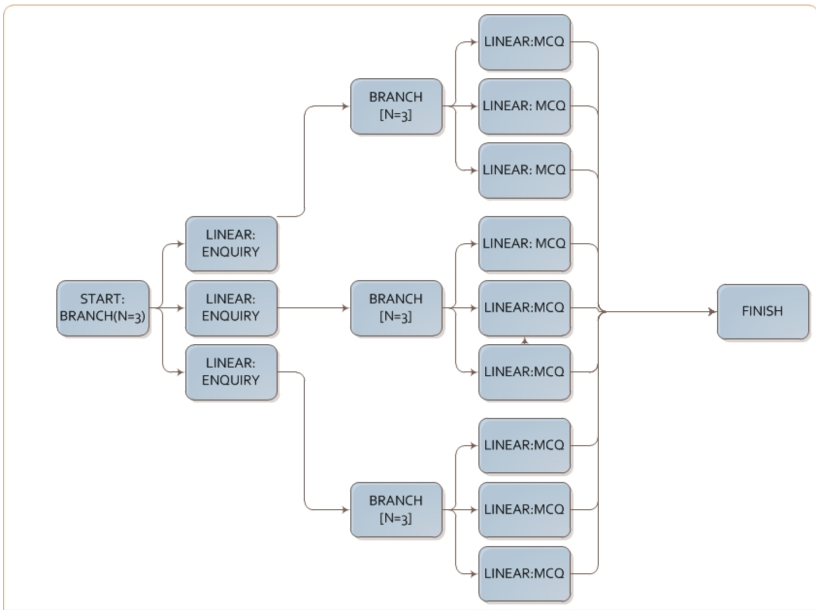
Schematic of a pure linear case

b.)



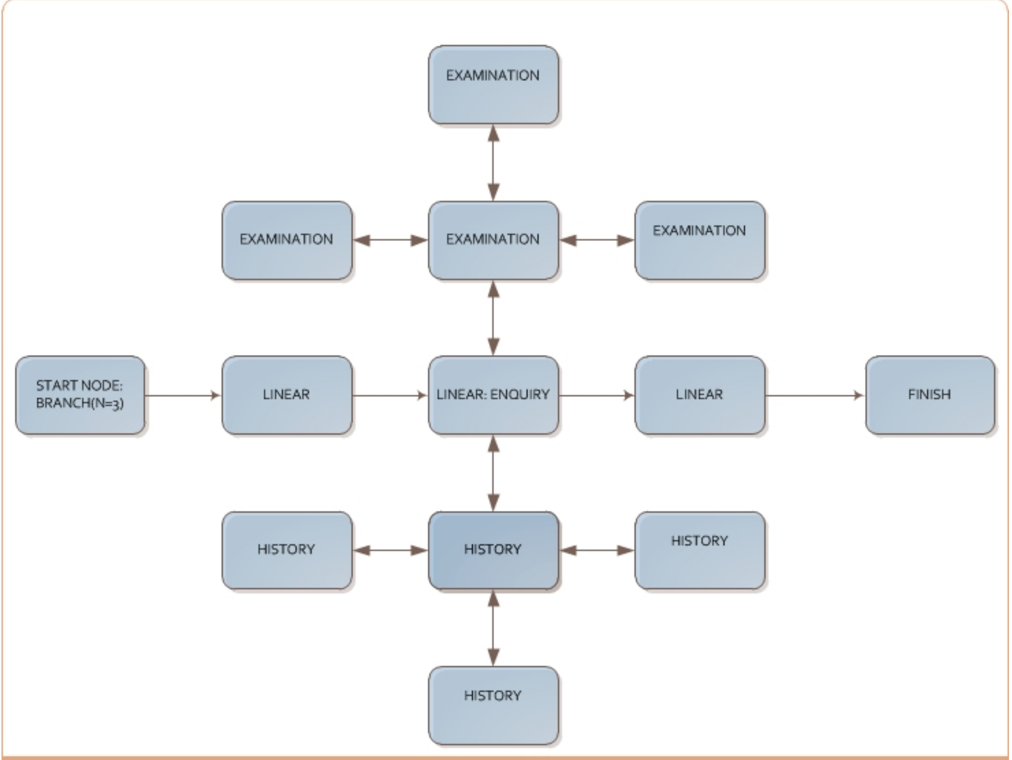
Schematic of two essentially linear cases that each incorporate a branch with three different decisions, however student will always follow the same path irrespective of the decision taken

c.)



Schematic of a four step VP Case where every user will encounter two branching nodes, and two continuation nodes, and the same final outcome.

Figure 4 A representation of examples of different VP pathways with linear (a) and branching cases (b,c)



Schematic of the 'wheel and spoke' method of creating authentic virtual patient cases, the arrows show the directions the learner can travel

Figure 5 My representation of a wheel and spoke VP design, with a core of linear structure but with vertical spokes travelling upwards and downwards

1.4. VPs in the e-learning and wider educational literature

Evidence from the e-learning literature, and case based learning literature may be helpful. Best-practice evidence from e-learning and instructional design (Clark and Mayer, 2008) is potentially applicable to VPs (Cook and Triola, 2009). Consensus does exist, the Cambridge handbook of Multimedia Learning (Mayer, 2005) discusses practical elements of e-learning design, and educational theory. Some principles are particularly relevant to VPs such as 'guided discovery learning theory' where students interrogate a VP depending on their own motivation and curiosity, selecting relevant questions (de Jong, 2005). These principles are helpful, and have yet to be evidenced in VPs (Cook et al, 2010). At a broader level still, these concepts can be considered within the established frameworks of adult learning theory in medical education (Green and Ellis, 1997), such as David Kolb's model of experiential learning (Kolb, 1984) or the pattern of reflective practice described by Schön (1987). VPs could be considered a form of experiential learning if designed effectively, allowing reflecting in a virtual case environment 'in action', or subsequently 'on action'. These are potential research areas for VPs.

The literature on case based learning, web based learning and VP integration is potentially informative. For example a review by Kim et al., (2006) suggests focussing on aspects such as relevant content, structure, purpose, relevance, realism, engagement, authentic, challenging, and instructional design. The literature on web based learning and its development (Cook, 2005a), and contemporary e-learning research (Conole et al., 2006).

1.5. VPs and Clinical Reasoning

VPs have been described as being best placed to teach clinical reasoning skills (Cook and Triola, 2009) as opposed to other competencies. Other researchers have endorsed this view (Cook et al., 2010, Forsberg et al., 2010, Ellaway and Davies, 2011, Huwendiek and de Leng, 2010). For this reason in this section I will review clinical reasoning, along with barriers to clinical reasoning research.

1.5.1. Clinical reasoning and virtual patients.

In the literature there is no universal definition of clinical reasoning. Newble (2000) described it as follows:

"The cognitive process by which the information contained in a clinical case is synthesized, integrated with the physician's knowledge and experience, and used to diagnose and manage the patient's problem".

Higgs and Jones (2008) define clinical reasoning as:

"Clinical reasoning is a context-dependent way of thinking and decision making in professional practice to guide practice actions. It involves the construction of narratives to make sense of the multiple factors and interests pertaining to the current reasoning task. It occurs within a set of problem spaces informed by the practitioner's unique frames of reference, workplace context and practice models as well as by the patient or clients contexts…." (Higgs and Jones, 2008) p.4

This definition has direct relevance to VPs, in terms of constructing a narrative to evaluate different case factors. Kassirer (2010) stresses the importance of probabilistic reasoning and continuous reappraisal of clinical information as a case progresses, using Bayesian reasoning (Herrle et al., 2011). Bayesian reasoning in medicine is defined by Gill et al. as follows:

"Bayesians interpret the test result not as a categorical probability of a false positive but as the degree to which a positive or negative result adjusts the probability of a given disease. In this way, the test acts as an opinion modifier, updating a prior probability of disease to generate a posterior probability." (Gill et al., 2005) p.1080.

Outside of this probabilistic approach, established cognitive theories describing clinical reasoning traditionally outline two different cognitive processes that overlap. These are analytical and intuitive reasoning processes (Croskerry, 2009a). Analytical reasoning was initially described as a hypothetico-deductive approach (Elstein et al., 1978), later framed around a individuals construction of a framework for different clinical problem's termed 'illness scripts' by Schmidt et al. (1990). The role of intuitive or unconscious inductive reasoning (Barrows and Feltovich, 1987) was built on by Brooks et al. (1991) postulating that subject specific experience replaced higher cognitive function. In a 'universal model of diagnostic reasoning' (Croskerry, 2009b) these intuitive and non-analytical process are described to run simultaneously. These theories are relevant to VPs. Realistic case representation in accordance with these theories should present the opportunity to foster, assess, and practice clinical reasoning. This has led to evidence suggesting structured clinical reasoning instruction can improve the clinical reasoning ability of undergraduate students, such as the SNAPPS approach (Wolpaw et al., 2009), and structured approaches to Bayesian reasoning problems (Sedlmeier and Gigerenzer. 2001). The SNAPPS approach is as follows: **S**ummarise briefly the history and findings; **N**arrow the differential to two or three possibilities, **A**nalyse the differential by comparing and contrasting, **P**robe by asking questions about uncertainties, **P**lan management for the patient, **S**elect a case-related issue for self study. Other notable research has shown that Bayesian reasoning instruction is poorly performed by practicing physicians, but can be taught using simple approaches (Sedlmeier and Gigerenzer, 2001).

These theories may also help us to understand barriers to good clinical reasoning, such as mind wandering.

1.5.2. Barriers to clinical reasoning: mind wandering

Clinical reasoning is prone to error, and is therefore the subject of research into strategies to mitigate poor clinical reasoning performance, of which VPs may present one modality. Mind wandering defined as 'task unrelated thoughts' is thought to be an important modifiable phenomenon in medical practice (Smallwood et al., 2011) that causes poor clinical reasoning. Smallwood and colleagues cite it as a potential area for intervention in medical research. Mind wandering has been estimated to occupying half our waking consciousness (Killingsworth and Gilbert, 2010). Mind wandering is commoner in individuals who suffer from burnout, fatigue and depression (Smallwood et al., 2011), features commonly found in medical trainees and professionals (Dyrbye et al., 2010, Bateman et al., 2011). Mind wandering is likely to be important in the clinical reasoning process itself (Smallwood et al., 20011) a process known to be prone to error (Elstein, 2009). Mind wandering can be measured (Dixon and Bortolussi, 2013), and also appears to influence simulated practical skills (Yanko and Spalek, 2012). We know attention and participation in e-learning tasks using VPs is sometimes suboptimal (Edelbring et al., 2012). Strategies do exist to mitigate the impact of mind wandering such as 'mindfulness' (Ludwig and Kabat-Zinn, 2008). The seven properties described for mindfulness: patience, openness, trust, non-judging, non-striving, acceptance, and letting-go. Their use in clinical practice for physicians and patients is becoming increasingly topical (Sanyer and Fortenberry, 2013). Research in VPs with an awareness of mind wandering may identify new strategies for its measurement, frequency, predictive value for decision making, and mitigation of its effects.

Cook and Triola's (2009) description of a continuum of competencies for medical undergraduates is one description of the *role* of VPs. Logically VP design should also be informed by other educational theory. Other more contemporary theories are relevant to authoring, such as Mayer's theory of multimedia learning (Mayer, 2010) and have implications for VP design relating to processing of different media types. These are not discussed in detail in this document.

1.6. The Unanswered Questions of VP research.

These research efforts change the fundamental questions of VP research from *if*, to *how* contemporary VPs should be designed, used, delivered and implemented. The single published systematic literature review on computerised learning with VPs (Cook et al., 2010) did attempt a meta-analysis with the primary goal of measuring the benefit of a 'VP' intervention in terms of no intervention, and then via alternative computer assisted instructional techniques. The authors conclude:

> *"We believe that theory-based comparisons between different virtual patient designs, and rigorous qualitative studies, will clarify how to effectively use VPs for training health professionals."*

I have highlighted problems in research description, methods, reporting, and transparency, alongside important research in an ever-changing field. Finally, we do not know how the inclusion of validated measurements of clinical reasoning should best be integrated into VP design, or how other questioning metrics such the number of choices to branch to. VPs today are a widely used educational resource. Hardware and software advances and uptake have resulted in a paradigm shift in what constitutes a VP. Although extensive work supports VP use, the individual design is under-researched, and this study intends to address calls in the literature for further VP research.

Section 2. Research: Planning, paradigms, practical choices

In the preceding section I have presented research evidence into the design, delivery and adoption of VPs. This section presents the original planning of this research within the context of a 2009 application to Arthritis Research UK for funding of this research project. I present the methodological frameworks used for this research, and introduce the qualitative and quantitative research methods used in this thesis.

2.1. Planning

The research was initially planned in 2009 with my project application to an education research funder, Arthritis Research UK. I was the principle investigator in the research grant application, supported by my PhD thesis supervisors based at the University of Warwick and University Hospitals Coventry and Warwickshire NHS trust. Funding was awarded in July 2010, grant number 19330 (Arthritis Research UK, 2010), with the project beginning August 2010. The original ten-stage research plan submitted to the research funder is shown below (Figure 6). The original timeline of the research plan from the funding proposal can be seen in Figure 7.

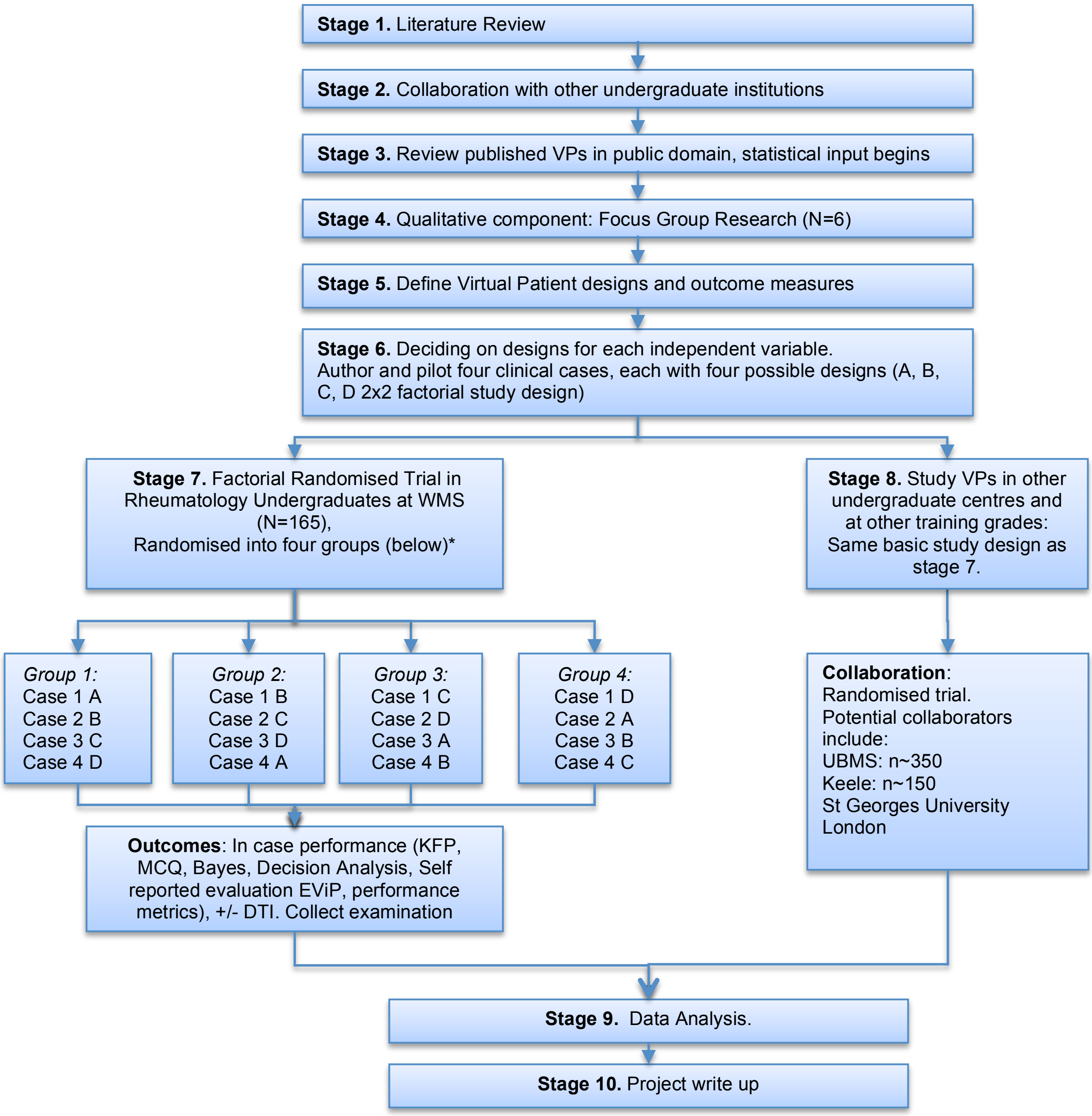


Figure 6 The original 10-stage research plan for the three year study component protocol.

	Pre commencement of Study	0- 2	3-4	5-6	7-8	9-10	11-12	13-14	15-16	17-18	19-20	21-22	23-24	25-26	27-28	29-30	31-32	33-34	35-36
Ethical Approval: Written approval from the NHS LREC as 'educational research' outside remit of NHS obtained August 2009																			
Stage 1: Literature review																			
Stage 2. Collaboration with local universities																			
Stage 3. Review published virtual patients, statistical consultation																			
Stage 4. Qualitative study of students attitudes to VP structure and feedback																			
Stage 5. Define Virtual Patient Designs, and outcome measures																			
Stagd 6. Create and Pilot VP Cases																			
Stage 7. Prospective randomised trial of undergraduate medical students at Warwick Medical School																			
Stage 8. Collaboration with other centres, and foundation year doctors																			
Stage 9. Data analysis																			
Stage 10. Finalising Project Write up																			
Projects that continue throughout study include liaison between departments and local universities, with IT services, statistical input, ongoing literature review, professional development																			

Figure 7 A reproduction of the original timeline for the research, submitted prior to the research being funded, produced in June 2010.

2.2. Paradigms: Ontological and Epistemological perspectives

This research planning of qualitative and quantitative components draws from philosophical frameworks that reflect my own undergraduate and postgraduate training, relying predominantly on post-positivist and constructivist paradigms. In this section I will discuss the meaning and definitions of these concepts.

The qualitative and quantitative components have two different research paradigms, or philosophical perspectives. A research paradigm can be defined as follows and the term is analogous to epistemological stance (Bergman et al., 2012):

“A philosophical framework that underlies and affects research activities. What are the assumptions underlying one’s views on reality and knowledge? Synonyms: theoretical or epistemological stance, world view” (Bergman et al., 2012), p. 591

My medical training follows a traditionally positivist and post-positivist viewpoint (Goldenberg, 2006). My postgraduate training in medical education introduced the importance of different research paradigms. Many research frameworks have been applied to medical education research: positivism, post-positivism, critical theory, interpretivism and constructivism (Bunnis and Kelly, 2010). Each paradigm has its own ontology (view of reality), epistemology (origins and character of knowledge about the reality) and methodology (dictated and planned in deliberately to address the research question). The two main research paradigms, used for this work are post-positivism and constructivism, and in this section I justify adopting conflicting , but in my view potentially complementary perspectives. O’Brien (1993) described these theoretical perspectives as akin to a kaleidoscope, with the different lenses representing different perspectives that may influence the methodology: for example study design and sampling.

2.2.1. Use of quantitative and qualitative research design

I have designed research with qualitative and quantitative components broadly from a post-positivist and constructivist viewpoint. Positivism is attributed to its founder Auguste Comte (1798-1857) and the works of Francis Bacon. A positivist view would suggest scientific method, and rigour can be used to explain why a particular educational intervention, such as a VP design, would be more effective. From an ontological perspective, positivism is realism. A ‘reality’ exists, i.e. there is an optimal effective VP design for a particular circumstance. It is not the individual who creates that reality. Experimental design would rely on hypothesis generation, testing, with a goal of predicting how VPs should be used and designed in education. This is an objectivist epistemology, with a single reality, contrasting to other theoretical perspectives such as constructivism. Epistomology has been defined as:

> *“Theory of knowledge. What are the origin, nature, and limits of knowledge about reality?” (Bergman et al., 2012)*

Positivism places less value on opinion, emotions, beliefs or impressions, and is more interested in scientific cause and effect. Clearly, particularly in education research, this potentially idealistic view of the world has evolved to be replaced by post-positivism, proposed by Kuhn (1922-1996).

Ontology is defined as “the nature of reality” (Bunnis and Kelly, 2009). This is my own perspective for the quantitative component of this work, where the ontology is described as ‘critical realism’ (Illing, 2010), or objectivism (Bergman et al. 2012) where “reality is static or fixed” (Bunnis and Kelly, 2009) and is observable. Post positivism’s ontology of critical realism/ critical objectivism describes the limitations from research methods to evidence a reality, due to the complexities of the subject of enquiry. Post-positivism suggests that because of these complexities, reality is not objectively identifiable. Like positivism, the epistemology is objectivism, and a focus on the scientific method. Here it is the validity, reliability, research method and the reproducibility of findings that moves the emphasis from proof to probability, and hypothesis testing and deduction. An example of this would be recurrent experiments in different areas that fail to disprove a hypothesis as being evidence that the hypothesis is true. One example of how this post positivist viewpoint has shaped this research is the importance I have placed on data triangulation in my research findings. This is of practical importance, because not all schools of GT would use triangulation based on different research paradigms.

2.2.2. Grounded theory: schools, paradigms, and a personal stance

Grounded theory (GT) has its origins in positivism. In fact, the original text title reflects the positivist grounding: a 'discovery of Grounded Theory' (Glaser and Strauss, 1967). GT has evolved to have different schools of GT, each with their own ontological and epistemological perspectives. It is logical to present GT to the wider educational research audience as fitting into one of these paradigms, for example as an 'interpretivist' (Bunnis and Kelly, 2010). In fact there are three main iterations, or schools of GT: Corbin's School (Corbin and Strauss, 2008); Constructivist GT supported by Kathy Charmaz (2006); and "Glaserian" GT (Stern, 1995, Glaser, 1978). I have used a school of GT described by Juliet Corbin, and adopt a similar epistemological stance. I describe GT in more detail in Section 3, p.67, however, I present a description of GT here to help explain my research stance.

Corbin herself does not feel that her school of GT fits within any single research paradigm, but acknowledges the post-positivist origins and constructivist elements to the methods she describes. In planning this research I have embraced this perspective, using both constructivist and post-positivist views. My post-positivist stance embraces use of quantitative data to support findings (triangulation). From a constructionist perspective, GT can provide analysis of case design that is lacking from a controlled experiment. A constructivist view suggests no single reality exists, and the ontology is relativism. Relativism is defined as *"reality is socially and experientially based; multiple realities exist, change, conflict and/or become more crystallised*" (p.545, Bergman et al., 2012). By embracing these elements I can explore reality depending on social interactions and experiences that may conflict, compete and evolve. Constructivism allows me to view VPs as dependent on the context and social circumstances they are used in. I believe it is likely that no single ideal VP design blueprint exists; the 'perfect' approach does not exist and cannot be discovered. Instead how different designs impact learners will be different for different individuals in different social and cultural circumstances. Embracing constructivist elements of GT allows the open understanding of how an individual constructs meaning from a VP. This may range from the contents of an e-learning package, to the context in which they sit the VP. The research challenge is to understand how these elements influence meaning, authenticity and experience. Despite this from a post positivist perspective as an author of VPs, I believe in the potential for evidence based experimental evidence to support suggestions rules and structures for VP design. The main quantitative component reflects the positivist perspectives. In an experimental controlled study I test the hypothesis that VP design will influence learning and user experience when using a VP.

2.2.3. Reconciling the use of different research paradigms

The research paradigm has dictated my choice of qualitative and quantitative methodologies for this research. I approached the project on the premise that by restricting my research methods to the use of one paradigm, or method would be a mistake, as it could discount either quantitative methods, or some qualitative approaches. The dogma surrounding the philosophical standpoints in for different *qualitative* research methods is now perhaps remitting, as described by Patton (2002), quoted by (Illing, 2010):

> *"Signs of détente and pragmatism now abound. Methodological tolerance, flexibility, eclecticism and concern for appropriateness rather than orthodoxy now characterise the practice literature and discussions of evaluation. Several developments seem to me to explain the withering of the methodological paradigms debate." (Patton, 2002)*

Patton's views are shared by Janice Morse who describes the easing of the disagreements between different schools of GT, a qualitative research method in "Tussles, Tensions and Resolutions" (Morse, 2009).

In summary my stance is analogous to the stance of Corbin (Corbin and Strauss, 2008), with positivist and constructivist elements shaping my choice of methods. In particular my preference for a flexible method to investigate how design influences the learning experiences students have with VPs dictated GT as my vehicle for the qualitative component.

2.3. Practical Choices: Five research principles

I have described the planning of a research study (2.1), the research paradigms used (2.2). I now discuss the practical decisions and ethos that underpinned my research proposal. This proposal planned close collaboration between a number of institutions to achieve the necessary student recruitment. I adopted five principles that feature consistently in the research grant application, this research, and the open publication of the research findings (Arthritis Research UK, 2010), shown in Table 2.

Research Principle	*Practical impact*	*Relevance*
1. Produce open access cases, and can be used, shared, edited under a creative commons attribution share alike licence. (http://creativecommons.org/). Wherever possible modifiable collaborative open access resources should be used for the research process to minimise costs.	Adopt open technology and avoid inflexible and expensive video resources.	Open access publication movement
2. Design VPs without specialist information technology input (Bateman and Davies, 2011), compliant with the MedBiquitous 2010 XML standard for sharing, cases use the MedBiquitous design possibilities such as branching cases (Huwendiek, 2009a).	Do not use natural language processing or bespoke features.	Relevance and generalisability
3. The research study must fit in scope and cost within a three-year education research fellowship.	Choice of materials	Practicality
4. The research should promote collaboration between centres in locally and nationally by presenting free educational resources.	Flexible research plan to fit within local curricula	Promotion of use collaboration and uptake
5. Use software with the ability to track unique users, data tracking of multiple metrics to evaluate decisions and performance within the environment (Cook et. al. 2010).	Choice of software	Originality and impact.

Table 2 The five principles for the research

Based on an options appraisal of the flexibility, immediate and long term costs, sustainability, functionality, and risks associated with each of the development packages, I chose the software VP-Sim/ DecisionSim (DecisionSim-LLC, 2012) to conduct the research. DecisionSim was created originally by the University of Pittsburgh and is an iteration of the software VP-Sim developed in Pittsburgh

(McGee et al., 1998). It is compatible with all of the principles of the research outlined in Table 2 (tracking, branching, open publication, MedBiquitous compliant).

Section 3. Review of Qualitative Research Methods

In the previous section I have introduced the VP research application with an initial quantitative research component. In this section I describe my choice of grounded theory (GT) as my choice of qualitative research methodology. A sometimes controversial method, this section will debate the different schools of GT, and their practical impact on research planning, data collection, theoretical abstraction and interpretation.

3.1. Choice of GT

Qualitative research comprises of a number of distinct methodological approaches. In this section I discuss the choice of a qualitative research component, the selection of GT for study, the use of focus groups to conduct the qualitative research, and methodological limitations of the study.

Qualitative research is defined by Lingard and Kennedy (2010)as follows:

> *"Qualitative researchers study social, relational and experiential phenomenon in their natural setting… How and what questions are particularly suited to qualitative research. (Lingard and Kennedy, 2010), p.323*

Kuper et al. (2008b) describe the process, and materials used for qualitative research as follows:

> *"…aims to generate in depth accounts from individuals and groups by talking with them, watching their behaviour, and analysing their artefacts (such as diaries, meeting minutes, photographs) and taking into account the different contexts in which they are based." (Kuper et al., 2008a), p. 405*

My research is intended to answer the following question. How and why different design principles influence student learning experiences when using VPs? The goal is to build theory and understanding of student experiences with VPs that are based on real student experiences, but flexible enough to include appropriate data, and sample the appropriate population. Here the research question has shaped the research paradigm that I have used (and to an extent, *vice versa*), and the choice of GT as a methodology. I have adopted a GT approach initially described by Barney Glaser and Anselm Strauss (1967). In Section 2.2 I alluded to the different schools of GT (Morse, 2009). I have chosen the method described by Strauss and Corbin (1998b). They describe GT as:

> *"theory that was derived from the data, systematically gathered and analysed through the research process. In this method, data collection, analysis, and eventual theory stand in close relationship with one another. A researcher does not begin the project with a preconceived theory in mind" (Corbin and Strauss, 1998), p.8*

This definition emphasises the importance of properties in common to all schools of GT: theory grounded in the data, iterative study design, purposeful sampling, and comparison of data. The theory developed in the research should be ‘grounded’ in the data, and not clouded by the researchers own background or preconceptions. Although researchers have ideas as starting points for research, described as ‘sensitising concepts’ (Bowen, 2006), and researchers should be aware, or ‘reflexive’ to them. Reflexivity is defined as follows: (Malterud, 2001)

> *“An attitude of attending systematically to the context of knowledge construction, especially to the effect of the researcher, at every step of the research process” (Malterud, 2001), p.484*

3.2. GT is a family of research methods

GT is a qualitative research methodology first described by Glaser and Strauss (1967), two American Sociologists. Dr Juliet Corbin was a nursing researcher who worked with Anselm Strauss for over 16 years (Morse, 2009). GT has evolved to be an umbrella term for a family of research methodologies with a fundamental set of core principles, with individual research groups practicing using separate methodologies from slightly different epistemological perspectives (Charmaz, 2006, Morse, 2009). The fundamental components of GT, is the generation of theory, derived from and grounded in the data collected by the researcher. Glaser and Strauss railed against the need for predetermined hypotheses to be tested in GT:

> *"we are also trying, through this book, to strengthen the mandate for generating theory, to help provide a defence against doctrinaire approaches to verification” (Glaser and Strauss, 1967), p.7*

Glaser and Strauss describe a systematic approach for allowing data to emerge in a process of categorising findings, purposeful sampling, iteration, and theory generation. They encouraged authors:

> *"literally to ignore the literature of theory and fact on the area under study" (Glaser and Strauss, 1967), p.37*

Glaser and Strauss described an approach where data must first be broken down, fractured into manageable meaningful units, in a process called coding that allows categorisation. GT uses a constant comparison technique. This means new data is compared and contrasted with existing data, potentially dictating further data sampling (theoretical sampling). Data analysis and data collection occur concurrently. GT was described in the context of care giving and illness experiences (Glaser and Strauss, 1965), it is now perhaps the most widely used qualitative research method in the social sciences.

Glaser and Strauss distinguished GT from other qualitative data based on theoretical sampling, and constant comparative analysis. Data collection was described as completed when 'saturation' occurred, when no new themes emerged. Researchers own skills directed the process of 'what' to code and categorise, described as "theoretical sensitivity" by Glaser and Strauss. The data itself can be in any form: such as an interview, transcript, video, or focus group.

3.2.1. The Evolution of GT to several families

The flexibility incorporated into GT led to debate and discord. In the decade following the 1967 publication, after initial collaboration (Glaser and Strauss, 1973) the two researchers parted company, and developed separate schools of GT (Charmaz, 2004), initially the split described as 'Glaserian' and 'Strausserian' (Stern, 1995).

Anselm Strauss went on to publish GT methodology that involved the use of a research paradigm, and a more explicit set of instructions for researchers (Strauss, 1987). Juliet Corbin, research assistant to Strauss, then co-researcher, collaborated with Strauss on the first work in 1990 the "Basics of Grounded Theory". Following Strauss's death in 1996, Corbin further developed the text that underpins the methodology I use (Strauss and Corbin, 1990a, Strauss and Corbin, 1998b, Corbin and Strauss, 2008). In each Corbin describes a coding paradigm, a tool to direct the coding process. Corbin and Strauss define the paradigm in their most recent publication (2008) as:

> *"the paradigm is a perspective, a set of questions that can be applied to data to help the analyst draw out contextual factors and identify relationships between context and process." (Corbin and Strauss, 2008), p.89*

This research paradigm they describe revolves around the principles of conditions, interactions and emotions, and consequences of events. This paradigm is perhaps the simplest example of how the schools of GT disagree. Glaser publication 'Basics of

GT analysis: emergence vs. forcing' (Glaser, 1992), suggests the coding paradigm potentially forces data into different categories arguing an 'identikit' approach is against the fundamental principles of GT, and proposes a different coding approach. Both schools have relative merits acknowledged by leading GT proponents (Kelle, 2007).

In summary GT evolved into different schools from its origins, with the work of Glaser and Strauss approaching similar concepts in different ways, Strauss proposing a more structured approach to data interpretation, with its accompanying limitations.

3.3. Corbin's GT: a description of the method used in this research

For the remainder of this section I will principally discuss the GT methods proposed by Corbin (2008). I present Corbin's approach to coding, 'memoing', and coding paradigms.

Coding allows the qualitative researcher to analyse emergent data, allowing categorisation, and the abstraction of meaning from data in any form of collection. Gibbs (2007) defined coding as:

> *"Coding is how you define what the data you are analysing is about. It involves identifying and recording one or more passages of test or other data items such as the parts of pictures that in some sense, exemplify the same*

theoretical or descriptive ideas. Usually, several passages are identified and they are then linked with a name for that idea, the code." (Gibbs, 2007) p.39

More simply, codes have been defined as follows:

> *"Codes capture patterns and themes, and cluster them under an evocative title." (Lempert, 2007) p.248*

This does not answer what a researcher should code. Glaser and Strauss provide guidance in the introduction to their original work (1967), drawing on the concept of theoretical sensitivity.

> *"He must have a perspective that will help him see relevant data and abstract significant categories from his scrutiny of the data." (Glaser and Strauss, 1967) p.3*

Line by line coding is suggested by Corbin as an appropriate first step for GT as it means going through the transcript in full, with an analysis unbiased by the researchers own presuppositions. Terminology for codes categories and themes has been described as "rather mysterious" to the average reader (Gibbs, 2007), and reflect the split between Glaser and Strauss. For example, Barney Glaser talks of "theoretical codes, coding and coding families" (Glaser, 1978), which has contrasting approaches to the approach of 'open, axial and selective coding' of Strauss and Corbin which I have adopted for this work (discussed in Section 3.3.1). Other researchers use different terminology such as 'index' (Lewis and Ritchie, 2003). The most important principle is that GT is an iterative process, as new data emerge, it should be compared against other codes and emergent categories, to promote understanding. Codes should be analytical and theoretical, not simply descriptive. The individual labels for each code are decided by the author, who may choose to use "in vivo" codes (Glaser and Strauss, 1967). Here, the name concept, or code, has been taken directly from the text.

'Memoing' is the process of making notes (memos) allowing one or more researchers to record the principles behind the coding process. Memos provide a benchmark for the development of more detailed analytical processes later in the research project, assisting researchers to conduct rigorous methodological practice. They help the researcher to conceptualise information from raw data, potentially influencing coding, sampling, prior and future analysis. Both coding and memoing can be helped by computer assisted qualitative data analysis (CAQDAS).

'Constant comparison' of data is another hallmark of GT. Corbin and others (Bryant and Charmaz, 2007) suggest comparison can be facilitated using a number of strategies. These include exploring deliberate extremes of a single dimension, to prompt the researcher to consider different possibilities: the 'Flip Flop' technique (Strauss and Corbin, 1990b). They also suggest other approaches such as 'far out comparisons', and 'waving the red flag'. 'Far out comparisons' refer to comparing different phenomena that share similar characteristics. 'Waving the red flag' points researchers to be wary of phrases that state certainty such as 'never' or 'always', and prompts the enquiry into 'what if' the situation actually occurred. The central process of GT methodology is the interpretation and assignment of meaning to the actions of participants, reflecting and comparing meaning.

3.3.1. Corbin's Coding Paradigm:

Corbin describes an approach to coding that occurs in three steps that are not mutually exclusive. Gibbs (2007) summarises the steps as follows:

> *"(1). Open coding, where text is read reflectively to identify relevant categories. (2). Axial coding, where categories are refined, developed and related or interconnected. (3). Selective coding, where the 'core category' or central category that ties all the other categories in the theory together into a story, is identified and related to the other categories." (Gibbs, 2007), p.50*

The process begins with open coding, a methodology of producing a concept from a piece of data or text. In the process of open coding concepts are organised into discrete categories. It may be the initial open codes are more descriptive rather than analytical, but these can then be refined at a later step. Following on from this axial coding organises the individual codes around different categories in a hierarchical structure. 'Lower level' concepts can be logically grouped into a smaller number of higher-level concepts. The process of axial coding is described by Strauss and Corbin (1998a, p112) in the second edition of basics of qualitative research as:

> *"The process of relating categories to their subcategories, termed "axial" because coding occurs around the axis of a category, linking categories at the level of properties and dimensions". (Strauss and Corbin, 1998), p112*

In axial coding, a category must represent a clear set of dimensions that are inter-related. The categories themselves are mutually exclusive. For example audio files, video, and pictures could all be categorised as 'accessory media', and defined as an extra resource for teaching. All the members of that category must share the common features, but have different dimensions, such as file size from (1 Kilobyte to 10 megabytes) or the time intended to use the media (seconds to hours). Here we can see that as the dimensions of accessory media emerge, so does the understanding of the parent category. No data would be excluded on the basis of the lack of an appropriate category- a category would need to be created.

Finally in GT, the final approach consists of selective coding, where a central phenomena is identified by the virtue of its relationships with the other categories. Strauss and Corbin (1998b) suggest you identify a single one. Further iterations of individual codes, categories and their dimensions occur at this stage, using a analytical and theoretical perspective (Gibbs, 2007), described by Corbin and Strauss (2008) as 'integration'. They define this as:

> *"the process of linking categories around a core category and refining and trimming the resulting theoretical construction." (2008, p.263)*

Corbin and Strauss respond to Glaser's criticism that the paradigm is formulaic suggesting "*not to fixate on the specifics of the paradigm*" (2008, p.90), and let the basic principles of emergent data persist. As discussed in Section 2.2.3 (p.65), I agree that dogma in GT methodology or the epistemological origins is likely to be counter-productive. Some qualitative research experts have categorised Corbin's work as continuing "*in the post-positivist tradition*" (Kennedy and Lingard, 2006). Critically, Corbin herself states that she explicitly has avoided a prescribed epistemological stance. She states:

> *"I realise there is no one "reality" out there waiting to be discovered… I agree with the constructivist viewpoint that concepts and theories are constructed by researchers."   (Corbin and Strauss, 2008), p.10*

The dissociation of the epistemological stance from method is in itself controversial, but Corbin states:

> *"I do not have a simple term to describe the method presented here" (Corbin and Strauss, 2008), p.7*

Corbin does embrace some of the newer constructivist viewpoints on GT as proposed principally by Charmaz (2006). There remains disagreement in the literature as to the most appropriate form of GT to use in Medical Education research (Watling and Lingard, 2012, Morse, 2009).

3.3.2. Problems with Adopting GT

The term GT has been misused in healthcare. Catherine Pope and Nicholas Mays (2007, p.71) describe the problems as follows:

> *"Unfortunately, the term 'grounded theory' has often been misused as a synonym for any sort of qualitative 'analysis', and sometimes weak analysis at that. It is not uncommon for research papers to report using 'grounded theory', but without any sign of the elements described." (Pope and Mays, 2007), p.71*

GT interpretation can be influenced by reflexivity (Shacklock and Smyth, 1998), researcher preconceptions (Strauss and Corbin, 1990b), and the epistemological stance (Cresswell, 2007). Shacklock and Smyth (1998) describe the substance of reflexivity as relating to the acknowledgement of the researchers own limitations which relate to all that they study:

> *"As we see it, the process of reflexivity is an attempt to identify, do something about, and acknowledge the limitations of the research: its location, its subjects, its process, its theoretical context, its data, its analysis, and how accounts recognize that the construction of knowledge takes place in the world and not apart from it....To not acknowledge the interests implicit in a critical agenda for the research, or to assume value-free positions of neutrality, is to assume `an obscene and dishonest position'." (Shaylock and Smith, 1998), p.6-7*

I accept I have my own researcher preconceptions, experiences and reflexivity shape the ideas that underpin my research questions, described as 'sensitising concepts' (Jupp, 2006). Charmaz (2003) does not see this as a barrier, but acknowledges they may create problems:

> *Sensitizing concepts offer ways of seeing, organizing, and understanding experience; they are embedded in our disciplinary emphases and perspectival proclivities. Although sensitizing concepts may deepen perception, they provide starting points for building analysis, not ending points for evading it. We may use sensitizing concepts* ***only**** *as points of departure from which to study the data. (Charmaz, 2003),* p.259

*bold emphasis reproduced from original

One strategy is to explicitly identify these preconceptions (Lingard et al., 2008). I have illustrated some personal examples of these in Figure 8.

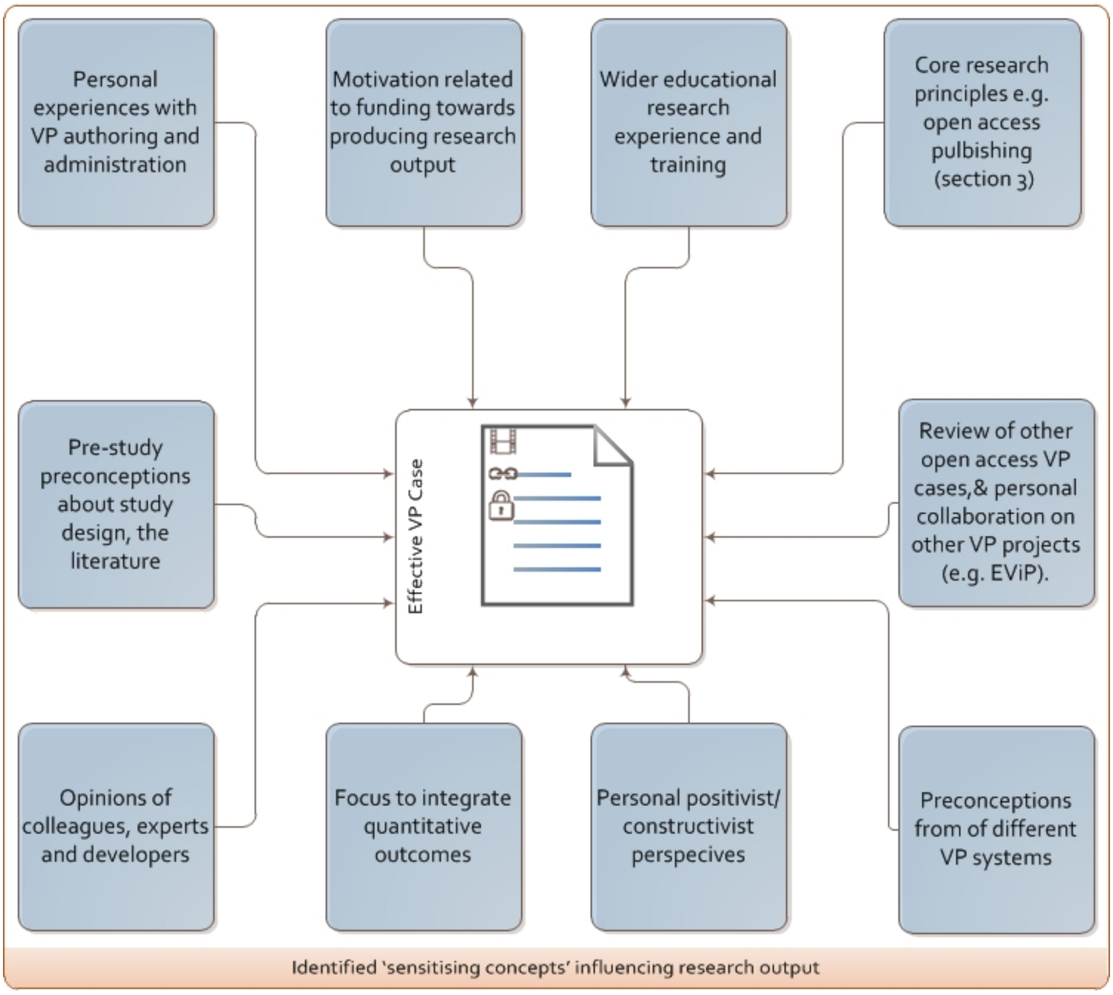


Figure 8 Examples of potentially overlapping sensitising concepts, preconceptions and reflexivity for this research

3.4. Focus Groups as a vehicle for research.

There are a number of approaches to gather information used in GT research. Options include semi-structured interviews, participant observation, and analysis of transcripts, each with relative strengths and weaknesses. Focus groups have been used in commercial market research, and by medical researchers for many years. They are suited to qualitative research methods and open questioning (Flick, 2002, Kitzinger, 1995). A focus groups is defined by Kruger and Casey (2000) as follows:

> *"A focus group study is a carefully planned series of discussions designed to obtain perceptions on a defined area of interest in a permissive, non-threatening environment. Each group is conducted with six to eight people by a skilled interviewer . . .Group members influence each other by responding to ideas and comments of others." (Kruger and Casey, 2000), p.5*

The skilled interviewer is typically called a moderator, a member of the research team who oversees the discussion. The moderator may also have an assistant during the focus group, termed a facilitator. The facilitator may also take on other roles determined by the researchers.

Focus groups have been used for research in musculoskeletal medicine (Pettersson et al., 2010), and in medical student attitudes towards both VPs (Botezatu et al., 2010a) and patient experiences (Kelly and Nisker, 2010). Focus groups have also been used to evaluate educational instructional design (Bridgemohan et al., 2005), and how trainee doctors learn in the workplace (Teunissen et al., 2007a).

The choice of a focus group as a method for this GT research is based on three principles. These are: (1) the ability to maximise the spread of participants sampled within a given time; (2) the appropriateness of the subject matter to maximise reflection within the group dynamic; (3) my own prior experience in designing and delivering focus groups. A focus group is likely in itself to provide a significant proportion (up to 70%) of the individual ideas elicited in by a series of individual interviews (Fern, 1982), making them an effective use of interview time. Focus groups allow exploration of group interaction and dynamics, described by Kitzinger (1995) as follows:

> *"When group dynamics work well the participants work alongside the researcher, taking the research in new and often unexpected directions." (Kitzinger, 1995), p.300*

As an exemplar I would not suggest using a focus group to study professionalism in medical education (Bateman et al., 2011), as the topic may itself influence the group dynamic.

3.4.1. Guiding Discussion in Focus Groups

Whilst there are several approaches to delivering focus groups, I have chosen to use an established methodology described by Krueger and Casey (2009). To help direct the discussion in a focus group, some focus groups use a defined set of topics, a 'discussion guide'. Henninck (2010) defines this as follows:

> *"A discussion guide is a pre-prepared list of discussion topics or actual questions used by a moderator to facilitate the group discussion." (Henninck, 2010), p.45*

I have used an iteration of this called the 'funnelled questioning route' (Krueger and Casey, 2000). A 'route' contains specific text and is read verbatim. This has a number of advantages. The 'funnel design' (Krueger and Casey, 2009) describes the introductory and opening questions, which move towards more specific questioning.

Kruger suggests five categories, these are: (1) these are the initial opening questions; (2) introductory questions; (3) transition questions; (4) key questions; (5) closing questions. Using a ‘funnelled’ questioning route, it allows the piloting of exact focus group questions prior to data collection, and standardised question delivery. The route, and questions can be revised either systematically (a ‘rolling discussion guide’), or through selective refinements when judged necessary (Stewart and Shamdasani, 1990).

3.4.2. Pitfalls with Focus Groups

There are a number of problems and pitfalls with focus groups. These include the moderator’s skills at promoting good group dynamics (Hennink, 2010). Individual students may dominate, or significant agreement may result in a subsequent paucity of depth or breadth of discussion. Researchers should be reflexive, aware of factors that may influence how participants behave. For example in this research I do not present personal ownership of the VPs used, this may promote ‘deference effects’, where students are reluctant to critically appraise the VPs.

In summary focus groups when used in an appropriate structured fashion can be a powerful research tool to support GT research.

Section 4. Qualitative Research. Virtual Patients: what works and why? A grounded theory study

In the opening three sections of this thesis I have outlined that VPs are an effective tool for teaching medical undergraduates. The design of VPs has been flagged as an important research question in undergraduate medical education. In this section I move from the introduction and planning of research to describe the major qualitative component of this thesis: a grounded theory focus group study into how VP design properties impact on undergraduate education.[1]

4.1. Research Question.

In the earlier sections I have highlighted the importance of Virtual Patients in Medical Education, the lack of evidence to support VP design, and the calls in the literature for further research. The research question is as follows:

> How and why do different design principles commonly used in VP development influence the experience undergraduate medical students?

4.2. Methods

This is a GT focus group study of medical undergraduates from one centre. I authored two thirty-minute VP cases for the explicit purpose of the research. I used the MedBiquitous VP standard described in the literature review (MedBiquitous Virtual Patient Working Group, 2010). I used the VP authoring software DecisionSim

---

[1] This qualitaitve research has been published as open access research.
Bateman, J., Allen, M., Samani, D., Kidd, J. & Davies, D. 2013. Virtual patient design: Exploring what works and why. A grounded theory study. *Medical Education,* 47, 595-606. PMID: 23662877,
For full text see appendix 8a, p.251.

V2.0 (DecisionSim-LLC, 2012). A schematic of the cases displaying the different pathways through the case and how the overall case structure is laid out is shown in Figure 9, p.87. The cases include design properties including linear, branching, and ‘wheel and spoke’ case narrative (Huwendiek et al., 2009a), created by combining different node types (Figure 1, p.24; Figure 4, p.50)

The other design principles are based on earlier sections of this thesis on my own literature review of virtual patients, other expert evidence (Cook and Triola, 2009), existing open access virtual patient cases, and experience in producing and authoring cases for a European Union research project, EViP (Smothers et al., 2008). The design principles studied are presented the ‘instructional design features’ in Table 3, p.88.

Figure 9 A schematic representing the overall case structure of the research cases. A detailed description of their components is found in table 3

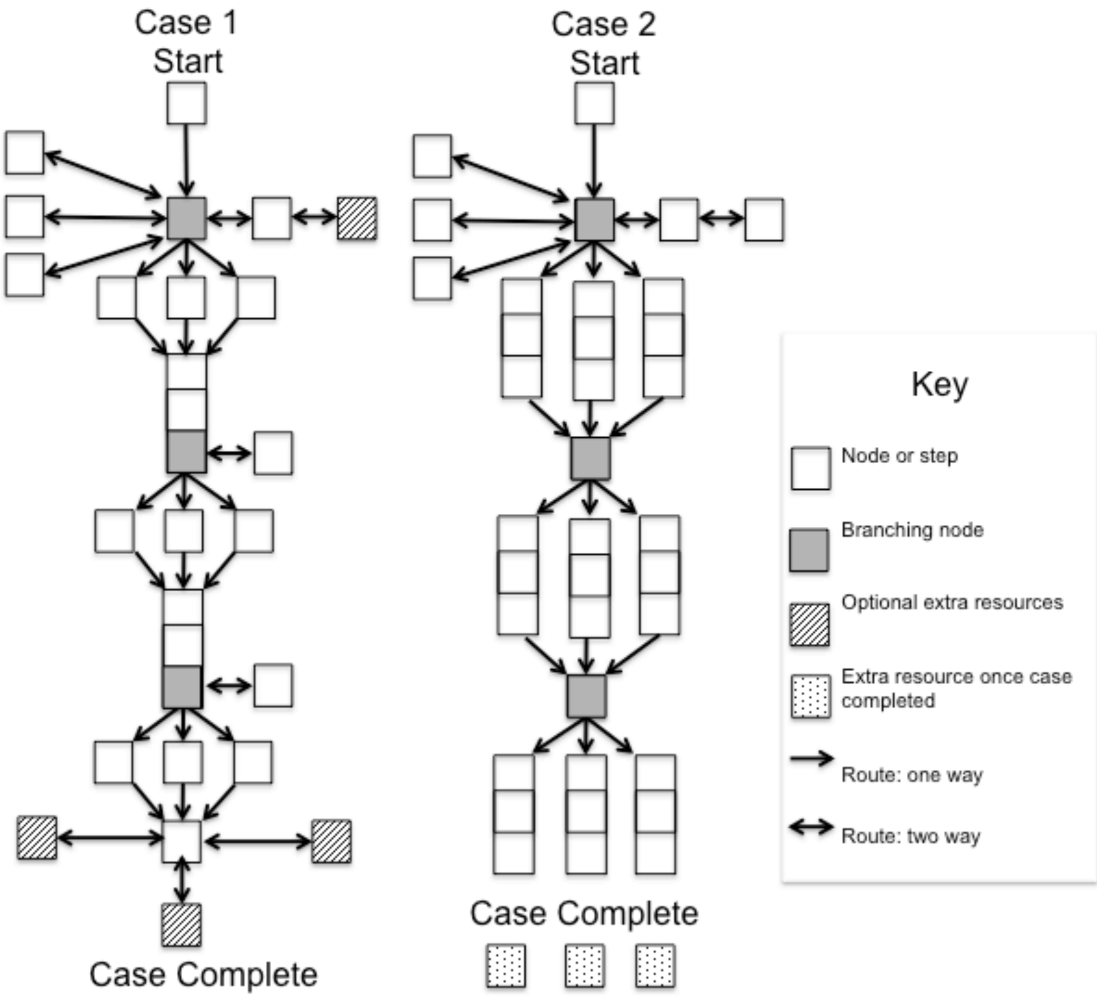


Table 3 The case design for the two research cases, showing case properties, supporting media, and instructional design features

Overview	***Research Case 1***	***Research Case 2***
Case Narrative	31-year-old Caucasian male, back pain, mechanical evolving to inflammatory sero-negative arthritis with knee effusion	27-year-old Asian female, arthralgia and compression neuropathy evolving to inflammatory arthritis (rheumatoid arthritis).
Final Diagnosis	Ankylosing Spondylitis, inflammatory knee arthritis	Rheumatoid Arthritis, carpal tunnel syndrome
Role of student sitting VP	Newly qualified doctor	Newly qualified doctor
Target Audience	Clinical Medical Students	Clinical Medical Students
Setting	Primary Case	Secondary Care
Number of characters in VP	3- Patient, General Practitioner, Rheumatologist	4- Patient, Orthopaedic surgeon, Physiotherapist, Rheumatologist
Time period Simulated	Months	Months
Principle Author	James Bateman	James Bateman
Authoring platform	DecisionSim V2.0	DecisionSim V2.0
Quality control	Peer review, piloting	Peer review, piloting
Case Properties		
Total Number of nodes (steps)	49	68
Case Type	Overall Linear, see S2 for further details	Overall Branched, see S2 for further details
Number of Branches	1	3
Substantial routes through VP	1	27 (3^3)
Choices at each branching point	3 (users redirected down linear path)	3
Minimal steps to complete VP	28	28
Node types	Question, branching, enquiry	Question, branching, enquiry
Time allocated by author	30 minutes	30 minutes
*Number of 'Wheel and spoke hubs**	3	2
Supporting Media		
Number of images in the case	24	16
Supporting images: Clinical	Physical Examination; Blood tests; GALS Screen, Annotated and normal Radiographs; Authentic reports (Radiography , microbiology, lab)	Physical Examination; Observation Chart; Blood tests; GALS Screen; Annotated and normal Radiographs; Patient HAQ **
Supporting images: Miscellaneous	Pictures of participants, environment, letterheads for correspondence, electronic results screens	Pictures of Patient, environment, letterheads for correspondence, electronic results screens
Audio files	No	No
Video files	No	No
Instructional Design Features Studied		
Structured promotion reasoning	No	Yes
History Taking: Enquiry chosen by student	Yes	Yes
History Taking: Information pre Given in authentic health care records	Variable through case	Variable through case
Investigation results: authentic presented	Yes, some	Yes, some
Investigation results: text only presented	Yes, some	Yes, some
Extra teaching resources within case	Yes	No
Extra teaching resources following case	No	Yes
Deliberate Errors included	No	Yes
Natural Language entry	No	No
Visual Signposting	Majority	Majority
Opportunity to question Peer	No	Yes
Common design Features Case 1&2		
In Case Questions and Assessment		
Bayes Reasoning, (n=1), Key feature problem style (10), Clinical decisions (3)		
Feedback in the cases		
Both given immediately following decisions and delayed to be presented authentically in case evolution. Includes why answers both right and wrong. Feedback given explicitly flagged as feedback, and tacit.		
Variable design features Case 1&2		
Visible Score (present vs. absent), Navigation (open navigation vs. closed), timer (present vs. absent)		

**Wheel and spoke refers to points where students an carry out different actions before carrying on down the linear pathway, for further information see Huwendiek S, et al.*[3] ***HAQ refers to the Health Assessment Questionnaire, a self reported patient questionnaire on wellbeing.*

Note: this table is reproduced from Bateman et al. (2013) Supplementary material

Two hospital consultant physicians, two general practitioners, and two doctors in specialist medical training piloted the cases. As part of the piloting reviewers submitted free text comments to improve, and comment on logistical, technical and medical aspects for the VP cases.

4.2.1. Setting and Participants.

The setting was University Hospital Coventry and Warwickshire Clinical Sciences Building Library and teaching rooms, a familiar environment to the students. I used an iterative purposeful sampling technique. A short introduction was used to invite volunteers to take part in research in musculoskeletal medical education. Inclusion criteria were that the subjects were medical students at Warwick Medical School. There were no specific exclusion criteria, for example students were not excluded on the basis of age, gender, or previous undergraduate degree programme. Volunteers were deliberately not informed that the research would be related to e-learning, but were aware that the method related to medical education, would require two hours of time, and would involve an activity and a focus group. To minimise any deference effects I did not acknowledge the personal ‘ownership’ or ‘authorship’ of the cases. Students who had volunteered to participate in the sessions were then given a participant information sheet, and had to sign informed consent. Written consent forms and participant information sheet (Figure 10) had received prior ethics approval.

Figure 10 Focus group participant information sheet

Qualitative Component Consent Form and Participant Information Sheet
Virtual Patients Focus Group (V1.4)

Name of Project: Virtual Patients Study

Researcher: Dr [redacted] **Supervisor:** [redacted]

Dear Medical Student,

I would like to invite you to take part in a focus group for an educational research study funded by Arthritis Research UK taking place at [redacted] Medical School. Before you decide on taking part in the focus group you need to understand why the research is being done and what it would involve for you. Please take time to read the following information carefully. Please ask if there is anything that is not clear or if you would like more information about the focus group before taking part. Please take your time in making the decision to take part. Some questions below will help you decide.

What is the purpose of the study?
We're interested in finding out the best way to design online virtual patients. We would value the input of some medical students on their opinions about the design properties of virtual patients. To do this we would ask you to try some online patients, and then take part in a focus group.

Why have I been invited?
We are inviting a group of medical students to take part.

Do I have to take part?
Your participation is completely voluntary.

What will taking part involve for me?
You will complete two-four virtual patient cases online. This will involve making decisions about a simulated patient case. It will be online. You will get feedback from your decisions. Virtual patients are not a test, and your performance is not relevant to the focus group, however your opinions on the patients will be discussed in a group setting.

The focus group would involve attending a short session (1hour) where a group facilitator (James Bateman) would introduce everyone and then ask a series of short questions to be debated by the group. As is routine in such groups, a recording of the group will be made electronically for future analysis. Individual comments could be noted and transcribed, but each individual's responses would be anonymised. The information will be then used by the researchers to help inform the development of some of the cases.

What are the possible benefits of taking part?
If you have not taken part in a focus group before you may be interested in seeing first hand what this involves as part of qualitative research. You will have the opportunity to work through some musculoskeletal medicine cases, which will hopefully be of educational value to you. In other studies students have enjoyed using virtual patients, and found them helpful.

Are there any risks in taking part?
The patients will take up a period of time. The facilitator should ensure that no serious arguments occur during the process. The study has been reviewed by the [redacted] Medical School Biomedical Research and Ethics Committee

What should I do if I am willing to take part?
You will be asked to sign a consent form detailing the above formally. You can withdraw from the focus group at any time.

Thank you for taking the time to read this information sheet.

4.2.1. Focus Group Funnelled questioning route.

I used a pre-planned funnelled questioning route. We piloted this focus group within our research group. This is shown below in Figure 11. For the purposes of this study, question design and structure has been calculated using available established literature, providing a series of key research questions over a one-hour period, with 50% of the time period spent discussing the key questions (Krueger and Casey, 2009). Question design is uni-dimensional and conversational therefore avoiding dichotomy when exploring areas. Our opening question ensures that each group member contributes to the discussion. The introductory and transition questions lead the students through introducing VPs as a concept, then, VP design. The styles of questions are conversational, and try not to bias the answers. For example 'were there any aspects of feedback that you liked or did not like, and why' rather than 'did you like x'. The five questions explore branching in VP cases, clinical reasoning, question styles, feedback, and any other issues. Our closing question explores a single design element they enjoyed, and then a single area for improvement.

Figure 11 Funnelled questioning route used in focus group research (Bateman et al., 2013)

OPENING QUESTION. Response from every student needed [00:00- 05:00]
[1] Welcome to the focus group. Perhaps we can all go round and introduce ourselves, perhaps your name, and a tell me interesting you have seen in your last clinical rotation.

INTRODUCTARY QUESTION [05:00- 10:00]
[2] The two cases you have sat are "Virtual patients", and teachers think use for them will be to in different ways. Have any of you used anything like this before?

[optional] Do you have any comments general on the cases?
What did if anything did you learn from them?

TRANSITION QUESTION [15:00]
[3]Like many areas of education, there's little evidence on how best to design and structure virtual patients cases to teach a certain area and help students learn. Do you think it is important about how these virtual patient cases are designed?

KEY QUESTIONS [15:00- 50:00- 5 MINUTES PER QUESTION, 10 MINUTES LEEWAY]
[4] In each VP case your choices lead you down a certain pathway. Did you notice a difference between how this worked in the two cases?

[Optional] If so, do you think it made much difference?
If you noticed this, How and why did you think it affected how good the case was?
[Optional] One of the cases actually had quite linear path, despite the different choices you had, whereas in the other case you could go down a number of different routes. Reflecting on the cases, how, if at all, did this affect you when completing the case?

[5] We would like the cases to help you make and weight up clinical decisions: thinking about how you make a diagnosis or choose a treatment. Think back to the cases: in one, you were regularly prompted from the outset to think about, reflect on, sift through, and modify the possibilities. How do you think this approach worked, and for example would you change or improve the prompting in the case?

[6] In the cases, you were asked to answer lots of different styles of questions, and make decisions. Do you think how the questions were constructed was OK?

[7] You had a number of different types of feedback through the cases. Were there any aspects of the feedback that you liked or did not like, and why?

[8] There are lots of ways that these cases can be put together as you have started to see. With this in mind, was there anything else, good or bad that we haven't discussed that you thought was important?

CLOSE: Response from every student [50:00]

[9] Finally, I would like to ask each of you individually a closing question. I would like each of you to pick out one design feature of the VP that you liked, and why, and one improvement you would make to either of the cases. So perhaps "***first student"*** you could start:

4.2.1.1. Analysis

A GT approach was used providing a constant comparison analysis of the data collected. I used a 'line-by-line' analysis for the first focus group, using memoing to record my initial thoughts on the themes that were being developed. During my initial open coding over 100 codes initially emerged. Memo writing was used to a narrative tool to help theorise, and provide context to student discussions and help formulate understanding of data. Early and late speculative memos were used. Theoretical categories were developed, allowing the development of concepts and subsequent theoretical sampling. Coding was developed, modified, with 'in-vivo' codes delineating different concepts from the transcript data. Memoing was used to initiate theorising and conceptual development. The qualitative analysis tool 'NVIVO 9' (QSR-International-Ltd., 2011) was used for developing the coding framework, and for comparison and discussion with supervising researchers. The open coding process was revisited by the principle research supervisors at scheduled meetings (DD and MA) and at departmental education research meeting. A constant comparison technique was used to identify concepts and hypotheses, establish saturation of concepts to inform further data collection, and refine the 'best-fit' for underlying categories. Axial coding of the data was conducted to merge categories into a more meaningful structure, which produced an emergent central phenomenon or core category. Categories and themes during the open coding were modified, renamed, merged and removed with a constant comparison using data from subsequent focus groups to inform the analysis from earlier work. Memos were used to help maintain analytical ideas. Purposeful sampling was used, with the option to sample students from any year group. Initially I chose students from year four, but as themes relating to the potential positive and negative impact on the year of student training recurred repeatedly across the groups, I purposefully sampled medical students when they began to have significant contact with patients (year 2).

Written feedback from the focus group facilitator (DS) was provided in electronic form following each focus group which included comments on the moderator's approach to the FG, and suggestions for the questioning route. I agreed the coding regimes with supervisors following meetings to discuss and agree coding of the transcript. Axial coding was followed by selective coding where concepts were linked to a central category that linked all of the other conceptual categories in the axial coding process. I used the coding paradigm suggested by Strauss and Corbin (1998, 2008). I then proceeded to selective coding, and the conceptual abstraction from the core themes to produce a pictorial model. This process continued in a cyclical iterative process to where the primary researcher (JB) developed constructed theory based on analysis of the paradigm. The theory was checked against memos and earlier categories, and reviewed and refined addressing gaps in logic. The model was presented at formal education review board meetings, and following a number of iterations and refinements a final model was produced.

4.3. Results

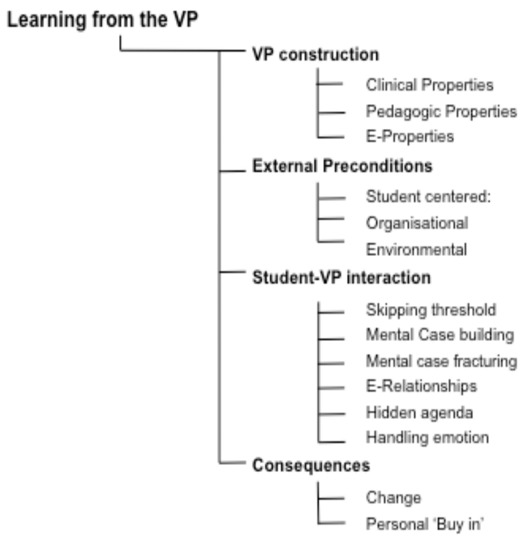
I conducted six focus groups in total. We invited 48 student volunteers to participate. On the day of the focus group, 46 attended (96% attendance, 29 males 27 females). The first four focus groups sampled medical students from year four. The final two focus groups sampled more junior medical students from year two.

Table 4 The dates and participants of focus groups in the qualitative research.

Focus Group	*Date*	*Participants attending (8 invited per focus group)*	*Year*	*Moderator*	*Facilitator*
1	25.2.11	8/8	**4**	JB	DS
2	11.3.11	8/8	**4**	JB	DS
3	23.3.11	8/8	**4**	JB	DS
4	24.3.11	7/8	**4**	JB	DS
5	5.4.11	7/8	**2**	JB	Not present
6	7.4.11	8	**2**	JB	DS

The analysis produced a core central phenomenon, ‘learning from the VP’. This had four subcategories (see Figure 12, right). These were: ‘VP construction’, ‘external preconditions’, ‘student-VP interaction’, and ‘consequences’. Each of these four subcategories are described in detail, and defined in the coming four sections. In each section I present quotations from students, with further quotations and definitions in the tables 5-8.

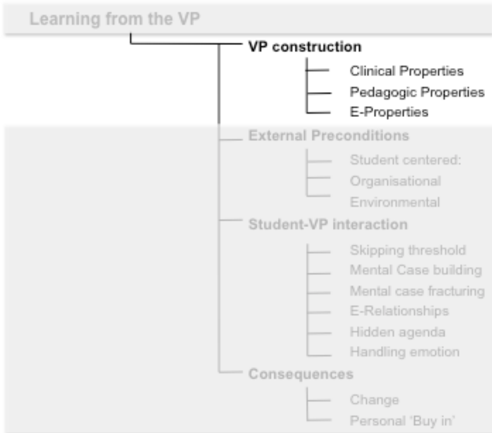
Figure 12 Schematic showing the central phenomenon, ‘learning from the VP’, and subcategories from the grounded theory analysis.



4.3.1. Category One: VP Construction

An overview of the category VP construction is shown in Table 5 and Table 6. VP construction has three component categories, 'Clinical properties', 'pedagogic properties' and 'E-properties'.

VP construction (right) describes the student's perspective on how they see the information and activities presented to them whilst completing the VP. These are grouped into three principle subcategories.



For the clinical properties of the case I have defined the two elements of clinical properties as: "*real life*" (*in-vivo* code), the clinical properties authored into a VP case and how they relate to actual clinical practice; and 'Pathway flux', how the flow of clinical and other information is presented between the student and the VP (see Table 5).

Table 5 Clinical properties of the VP

Clinical Properties. Clinical properties authored into the virtual patient by an author	
"Real Life" *The clinical properties authored into a VP case and how they relate to actual clinical practice*	
	Environment. Simulation of the clinical environment, for example GP having past healthcare records from a patient
	Authenticity. The authenticity of the narrative and supporting educational materials in the case
	Scope and content: The extent to which healthcare domains are explored by the case, such as clinical knowledge, professionalism, clinical reasoning, local healthcare policy, and health service structure.
Pathway Flux: How the flow of clinical and other information is presented between the student and the VP	
	Channels and dams. The degree of freedom given to the student over their actions, progression and the narrative in the case.
.	**Evolution-Evaluation**. The extent to which data and information is presented, reviewed and evaluated as the case progresses.
	Clinical Inertia: How case progression is resisted by the quantity, quality, completeness and relevance of pathways, data and activities that contribute to cognitive load, realism and difficulty

Clinical properties are properties that students see as being authored into the VP cases, that is factors that relate to how they mimic "real life" in terms of authenticity, scope and the clinical environment in which the VP is presented. The students also saw 'pathway flux' as an important component in presenting the flow of information from the student to the VP, and this relates to properties. These are shown in Table 6. The following quote describes the property of authenticity as a component of one of the in-vivo codes, 'real life'.

"I like the way it's based on the way we've been taught so far… you start with the history and you take a detailed history, and I like that it actually gave you the option of collecting that history from that patient. … it still followed the steps that you would take in a normal situation which is getting a clear history, a systems review included of a patient and a condition… definitely something that applies to ***real life**** *and definitely something that would be useful."* EA, FG6, Year 2 student

** '**real life**', an in-vivo code occurs 65 across the six focus groups, being used at least four times in each. Note the facilitator or moderator does not mention 'real life'.*

The following quote (below) is an example of a quote that codes for 'clinical inertia' and pathway flux.

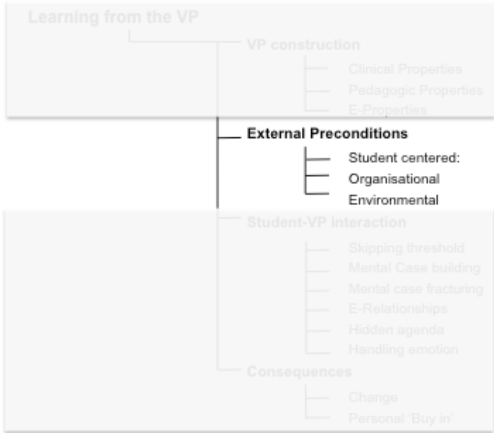
"the referral letter was good and bad, good because it's probably what we'd get, and bad because it was a bad referral letter… one of the questions was what is pertinent to this referral letter… and it had duration of symptoms, and you don't know how long its been going on for…" SR, FG3, Year 4 student

I present other examples, quotes and definitions for pedagogic properties and e-properties in Table 6, p.99.

Table 6 VP Construction: clinical, pedagogic, and electronic properties. In-vivo codes in italics

Pedagogic Properties: Teaching elements of VPs integrated into VP cases
Feedback: How feedback is delivered to the students as they complete a VP case. **Format effects.** Implications of different formats such as a letter, or a phone call, at different times through the case **Tailoring**: Extent to which student feedback is individualised, including comparisons with peer performance **Prompting reasoning:** Approaches that explicitly drive structured clinical reasoning. *"It was good to kind of think about the differentials… I do think the lack of knowledge was an influential factor, but it did help me question why is it that I'm including this one, and why is it that I'm including that one, I looked back to the history… you come across important factors… is that a long term condition, or is this acute... rule things out… I thought it was really good."* RR, FG5. **Decision Flux:** How decisions contribute to freedom to make decisions both correct and incorrect, and experience consequences of them. **Consequence effect:** Extent to which students feel their decisions impact further down the case narrative. **Limits and Forcing.** Being forced to undertake a particular action, decision, cognitive process or clinical experience irrespective of the apparent choices given. *"I quite liked the way that sometimes they got you to pick only three questions, which kind of got you maybe to think rather than ask just random questions. Think where your thoughts were going and what questions were important"* CD. FG6.
E-Properties: Electronic properties used provoking comment and outside of normal expectations for electronic interfaces.
E-Signposting: The helpful effect of signposting students using images of locations and particularly patients "I really liked on the first case the pictures. I know, I know it was just random adjudicators, but it kind of made you smile and if you've got that kind of visual stimulation, oh that's the GP OK, it kind of motivates you…" AR, FG1 **E-inertia:** Electronic properties authored into cases which slow or hinder a student interacting with a case. **Non e-tasks:** The use of items that don't require actions by the student for example summarising elements on paper **Software limits**: Desired software features from students, not present, which limit interaction. ***'Scroll scroll scroll':*** Impact of multimedia design including text format, length, steps, and image representation *"I think some of the pages were quite wordy, maybe it can be broken down into two instead of one, and squeezing all of the information into one page, it just gives me a headache "SS, FG2.*
E-Error: Electronic error as students sit a case, the cause of which may or may not be under the control of the case author

4.3.2. Category 2: External Preconditions



'External preconditions' are factors that relate both to the student and training organisation which are identified as being important to students when completing a VP. They appear to influence both the approach to the VP, and how useful it is. There are three subcategories of external preconditions. 'Student centred' describes factors that relate to the individual student and their personal motivations, goals and experience can influence and prejudice their learning from a VP. The following quote from focus group 4 describes a 'student goal' accepting an activity that resonates and aligns to an expected summative assessment target.

> *"I can see why its on there, because for finals, they are going to say, "what are your differentials, summarise the case in a sentence"* AR, FG1, Year 4. '**student goals'**

'Organisational elements' relate to the influence that an institution has when delivering any given VP, such as the background curriculum, or assessment strategy used at the school. Our students highlighted three elements: (1) the 'institution fingerprint'; (2) the 'assessed curriculum'; (3) 'local environmental elements'.

> *"I know we had an hour, and at the end people were leaving so I felt, sort of, hurry up."* DE, FG3, Year 4 student. **'environmental elements'**

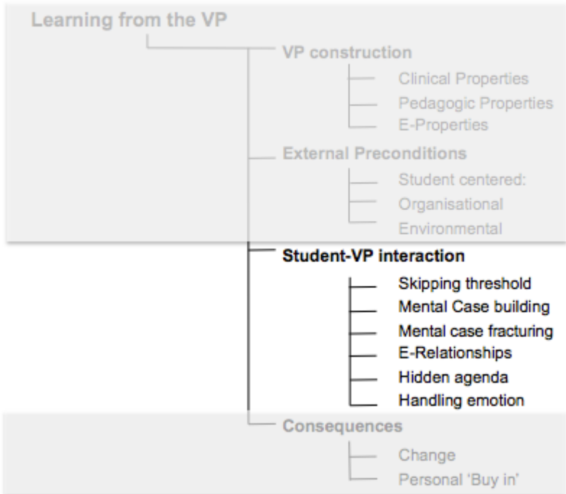
Further examples, explanations and definitions are shown below in Table 7.

Table 7. The category 'external preconditions'. Note: reproduced from Bateman et al. (2013). In press.

Student centred: *Student Centred factors that influences the utility of the VP case*	
	Electronic Prejudice: How both prior positive and negative e-learning experiences prejudice the approach to a VP
	Global Experience: The global knowledge and skills a student possesses about medical problems and healthcare systems
	Student Goals: The students personal goals for the learning activity, for example relating to assessment or professional development
Organisational Elements: *An organisations, educational aims, curriculum, assessment an evaluation of students*	
	Institution fingerprint: The organisational style and expectations of students when delivering and evaluating educational interventions. *"I don't know how that reflects on our teaching… we're quite often lead to a single best answer"* KG, FG1, Year 4 student.
	Assessed Curriculum: Factors that relate to the pedagogic, assessment and curricular approaches of an educational institution.
Environmental elements: *Local factors (location, computer hardware) that influence the how a student interacts with a VP case*	

4.3.3. Category 3: Student-VP interaction

The third category described is Student VP interaction. At its heart this describes the interplaying elements between a student and a VP, the relationship between the student and the case. There are six components of student VP interaction (see right, and Table 8, p.103).



This category is specific to each individual student, and is a function of the prior two categories 'VP construction and 'external preconditions'. The six sub-categories are as follows described in the following section.

Table 8. Student-VP interaction- the interactions between a student and the VP as they complete a case. Note: reproduced from Bateman et al. 2013.

STUDENT VP INTERACTION

Skipping threshold: A threshold, above which negative behaviour patterns occur, and interaction ceases to be constructive learning.

- **Efficient skipping:** Engagement is limited by a drive to efficiently pick out activities that are perceived to add value or be important
- **Judged credibility**: A constant appraisal of usefulness, quality, and interactivity of the VP to judge continuing with the case
- **Style:** Approaches where the style of questioning promoted lack of engagement and skipping.
  - "*there were some of the questions where it was you selected one answer, and it wouldn't allow you to go through until you clicked the right one…. At that point I'd given up and was just guessing, which obviously I would never do with a real patient."* DG FG5, year 2 student.

Mental Case building: Interactions which help the student to construct a mental representation of the case

- **Learning and assessment focus**: Added engagement resulting from perceived benefits in learning, future assessments, or the workplace
- **"*Thinking outside the list*": Strategies which encourage students to** think outside a predetermined list of answers in the task
- **Pathway growth.** Extent to which both decisions and branching pathways enhance students experiences, and experiential learning
  - *"I think I quite like the branching bit… because obviously in life there are a lot of different routes you can take and it doesn't necessarily mean one is… the best … I think its good to go a little bit off track…."* JC, FG2, Year 4 student.

Mental case fracturing: Interactions which impair the student constructing a mental representation of the case

- ***"Bogged Down"*** *Role* of ultimately irrelevant information either explicit (doctor suggesting behaviour for student) or tacit (information load)
- **Invisible Elephant.** Extent to which students see do not see feedback which is integrated into the case narrative, but not explicitly labelled.
- ***"Loss of control".*** Students perceived loss of control in the case that may or not be related to branching structures.
- **Pathway decay:** *Decay in learning, which occurs from, lost time, effort, uncertainty or motivation from being allowed to follow different routes.*
- **Contextual dissonance**: *Factors which clash with previous case assumptions, such as discordant information*
  - *"There were some discrepancies… the age on the GP records is different to when you are given the first stem, there is a 9 year difference, she was born in 1970 in one and 1979, I don't know whether or not that's relevant."* FH, FG1.
- **False Expectations:** Student false preconceptions about clinical scenarios and professional duties that are detrimental.
- **Points on the board**. Student primary focus becomes the assessment and scoring employed during the case.
- **E-failure:** A technical failure whose origins could be the students, author, VP software, or IT software, hardware and infrastructure

E-Relationships: How the students form relationships with electronic representations of patients and healthcare professionals.

- **Stereotyping:** Stereotyping or making moral professional or personal judgements about case participants."
  - FK. *"She was like 'oh I have a new partner, I want to start a new family'."*
  - AR. "*I was thinking, you've left it a bit late."* FG1, Year 4 students.
- **Relationship threshold**: Complexity of sustaining a more than two of relationships in a case (supervisors, patient, allied health professionals)

Hidden agenda: *Activity of deconstructing the VP and its components out (naturally, out of curiosity) or to improve performance*

- **Assessment subtext:** Interpreting the case in the context of the institution or teacher assessment strategies
  - MB*: "And when you think of it as an exam, you start looking for a style, because everyone has a style in the way they will write a question and answer, and you're trying to link the two up rather than thinking…"*
  - JC*: "What actually should I do?"* FG 2, year 4 students.
- **Case Template subtext**: Student devoting time to exploring real or perceived examiner VP design structure, for interest of to find patterns of assessment
  - *"It felt to me like essentially you went off for a little tangent for a couple of windows and then it would drop you back onto a common pathway towards to the end of whatever you did" JW: FG1, Year 4 student.*

Handling Emotion*: Students described the process of coping with different emotions during the case, eleven in total, listed here*

- Fairness, Humour, Comfort zone, Uncertainty, Fear, Confidence, Denial, Embarrassment, Pressure, Fatigue, Distraction.

The six categories of student VP interaction shown in Table 8, page 103. In the 'Skipping threshold' describes the concept of a point where a student disengages with the case narrative, and the personal reasons why students may behave in this way. In the table I demonstrate how a student that had described skipping inside the VP in the following quote, which I have coded as 'efficient skipping'.

JC. "Some of the pages had a lot of words on, and my eyes go, nah there's a lot to read there, and there's nothing to input, and I don't have to give anything, so therefore there's no need for me to read it because it wasn't about the case. That's just me being lazy…."

AA. "I think by nature we're all quite lazy, and we'll be like 'nah'."

JC. "Efficient." FG2, Year 4 students. '***Skipping: style'*** *and* ***'efficient skipping'***

This concept of a skipping threshold described by the students can be triangulated using time-stamped data records of student decisions within the cases (see Figure 13). Here student JC who describes skipping in the case can be seen to skip through a key step in the case, where the examination findings are presented. JC's time was 3 seconds on this node, the mean time for student colleagues was 30 seconds, mean reviewer time was 19 seconds.

a)

User [FG 2]	Seconds spent on node "Examination of Mrs Begum", case 2 of the qualitative research	Case Score
JC*	**3**	**13**
CB	40	18
JM	25	12
HD	33	16
AA	30	13
MB	52	12
AA	23	15
AM	33	12
Reviewer 1	23	N/A
Reviewer 2	15	N/A

b)

"I thought the use of lots of information from the history, test results and radiographs made it a very realistic case to work through." WA, Electronic free text comment, prior to FG3

"It was useful as I had to make the decision on what I would do, but if wrong I was redirected on the right course so I could learn from the case." JM, Electronic free text comment, prior to FG2

Figure 13 Two sources of data triangulation: a) shows empirical evidence from data logs of student JC 'skipping' through a node, compared to peers and reviewers. b) examples of free text comments from the self reported VP evaluation completed before the focus group.

I have described three separate reasons as to why students break through a skipping threshold: efficiency, that is completing an activity as quickly as possible; the credibility of the case, how much the case information and activities resonate with the students; and the characteristics of the learning material presented, for example block text requiring scrolling was not helpful.

'Mental case building' and 'mental case fracturing' are descriptions of how a mental representation of a VP can be enhanced or sabotaged. 'E-relationships' describe the formation and evolution of relationships between the student and various actors in the case, such as the patient, and other allied health professionals incorporated into the case. One component of 'e-relationships' is the 'relationship threshold'. This means that as the number of characters in the VP rises above three, students repeatedly had difficulties in maintaining relationships with them, producing confusion.

The ‘hidden agenda’ relates to student seeking alternative, and tacit components of the case that may relate to either an institution or author. Students described looking for elements such as traps and triggers that appear to mirror the students approach to summative assessment. Students described probing for patterns, styles of cases, or system vulnerabilities, which would allow them to work out the correct answer. There were no system vulnerabilities identified by the students but this did not stop the students from looking for them, as shown in the following quote.

> *MB: “And when you think of it as an exam*, you start looking for a style, because everyone has a style in the way they will write a question and answer, and you’re trying to link the two up rather than thinking…”*
>
> *JC: “What actually should I do?” FG 2, year 4 students.* ***‘hidden agenda’***
>
> *this also codes for ‘external preconditions’,

‘Handling emotions’ describes how a student experiences and deals with emotions during the interaction with the VP. Frequently the emotions relate to other healthcare professionals rather than the students with elements such as ‘embarrassment’ relating to the fear of working with ‘virtual’ colleagues. A student in the following quote links some of the emotion to the realism embodied by the case. Although logically some of the emotions such as fear, embarrassment appear to have negative connotations, the students regularly associated the emotions with more positive experiences.

*"It seemed quite realistic to me, like kind of both embarrassing and reassuring at the same time. Even though it was simulated, I did feel a bit embarrassed when I was being slightly corrected when I hadn't decided to refer the patient… That's…giving you the feedback in a realistic way, how it probably would happen in real life, <u>and I feel like I'm going to remember that a lot more because of that feeling of embarrassment</u>*****" AP FG5* ***, 'handling emotions'***

**The student here appears to link a negative emotion, embarrassment with a positive learning experience, that it was memorable to feel uncomfortable.*

4.3.4. Category Four: Consequences.

The fourth and final category was 'consequences' that is the impact on the students as a result of completing the case. There were two components of this, a description of a change in the students' knowledge, attitude or behaviour, and a change in a 'buy in' to VP cases in the future.

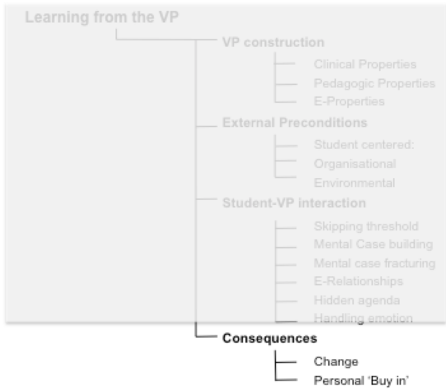


Table 9 Consequences- the results of a student engaging with an individual or series of VPs

CONSEQUENCES

Student Change: The impact on knowledge and behaviours in students future practice

Real World Reasoning: Incorporating processes taught in the VP that change clinical practice and approach to patients.

"I also liked the multiple choices question part, despite having 10 options for the blood results, what are the three most important ones…..It gets you into the mind-set of not ticking all of the boxes, which in theory you could probably do come August if you wanted to." BN, FG4, Year 4 student, [NOTE: August refers to the month graduates begin work as a qualified doctor]

Addressing weakness: Highlighting areas of knowledge skills or behaviours that are weak, and addressing those areas

"I kind of guessed the first one, and then I realised actually you can work it out… so it highlights your weaknesses I suppose" AR, FG1, year 4 student

Individualised Experience: Unique user learning experiences which depends on domains one to three.

"I have a different experience from you, again because I didn't look at the score…I stopped and I started doing the modified Schober's test… having that break where I didn't feel like I had to do anything, I was just learning." AL

Personal 'Buy In' Extent to which VP design influences current and future participation of VP cases.

Learning-realism Trade-off. An apparent trade off between learning and realism when faced with different design properties

"I think the question… is…. do you want learning or realism, because it was better to learn with the linear case, because obviously there's only one way to go with it… if you make the wrong decision…we're not going to learn what the right path necessarily is. Whereas its going to obviously be more realistic… so the second case was more realistic but the first one was a better learning experience." MB, FG2.

Future Uptake: The approach to voluntary or compulsory cases in future training

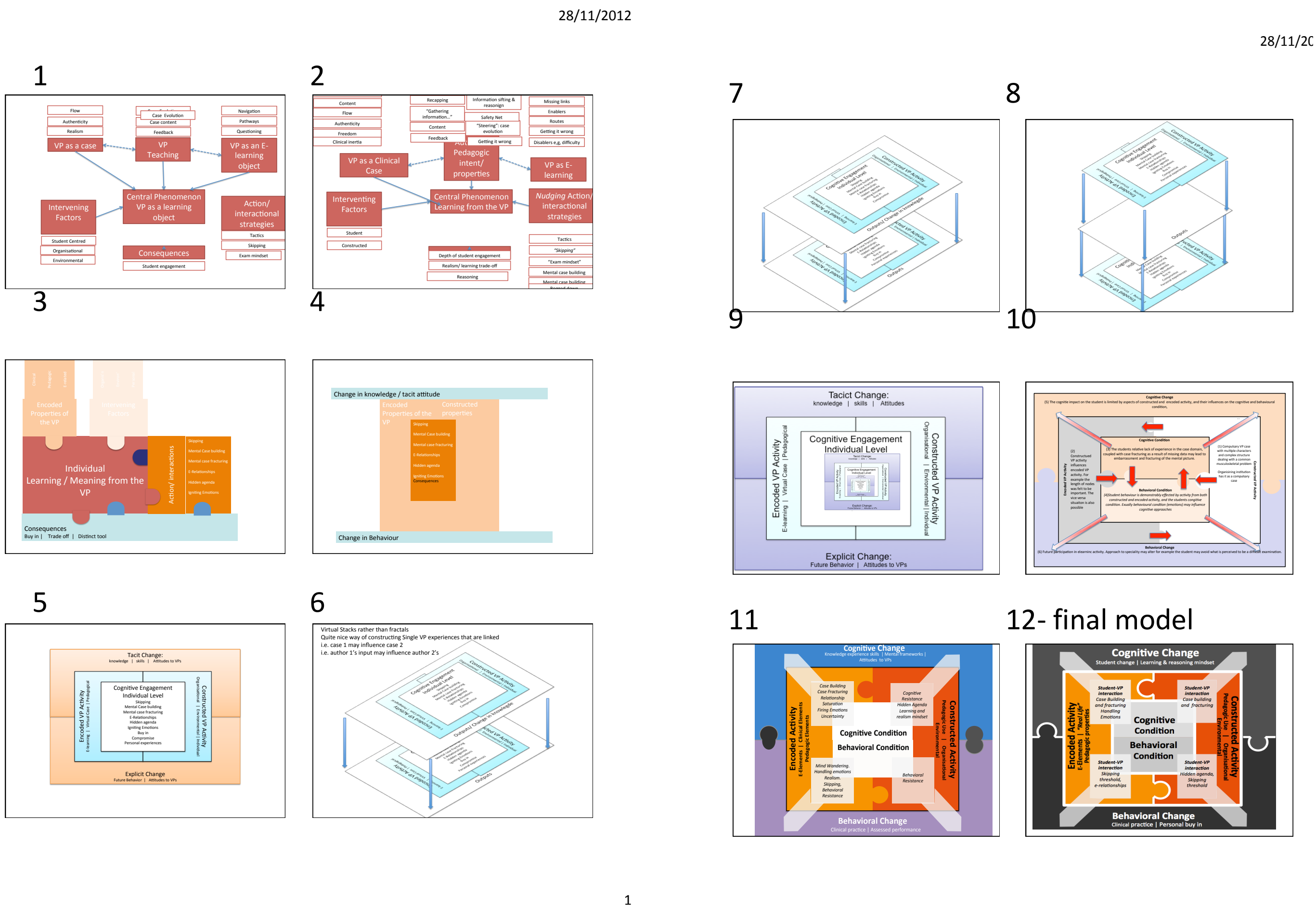
The following quote describes one student talking about how the case has impacted on practice.

> BN: *"I also liked the multiple choices question part, despite having 10 options for the blood results, what are the three most important ones... …It gets you into the mind-set of not ticking all of the boxes, which in theory you could probably do come August* if you wanted to."*
>
> FG4, Year 4 student, '**Real world reasoning'**.
>
> *August refers to the month graduates begin work as a qualified doctor

4.3.5. Developing the model.

The development of the model went through a series of steps, from an initial draft to the final model. For interest I present some of early iterations of the model in chronological order.

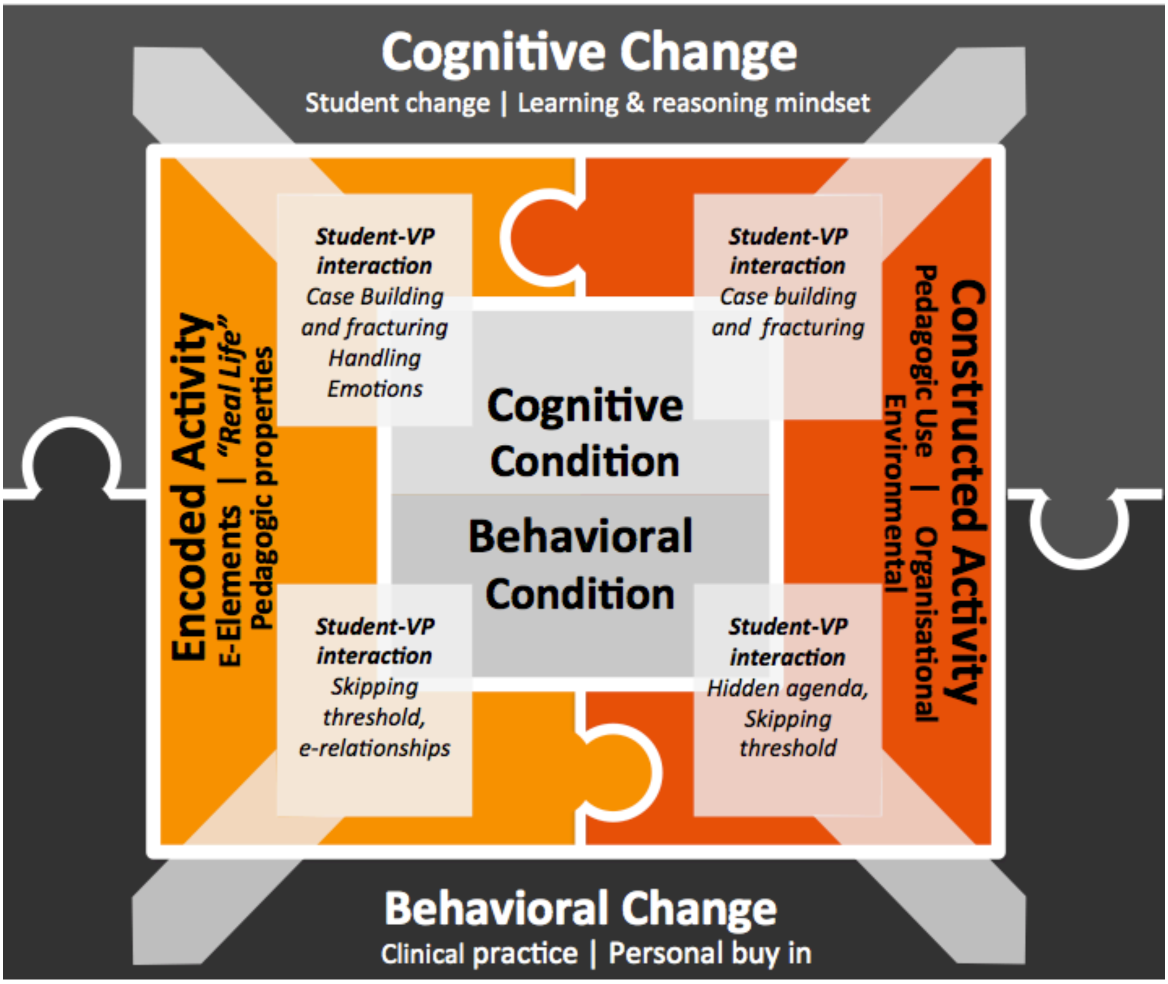
Figure 14 Examples of the abstraction of theory illustrates a progression of model development (see also appendix 3)



4.3.6. The model

The model is shown below in Figure 15. The model has three layers. In the inner layer I consider the student condition before they sit the VP. The middle layer comprises encoded and constructed activity. The outer layer represents student change, and there are cognitive and behavioural components. There are four boxes that overlap the inner two layers that influence student change, and this represents 'student-VP interaction'.

Figure 15 The virtual patient implementation model



4.3.7. Practical Application of the model: an example

To apply the model, I present a situation where another medical school, which already uses VP cases, wishes to integrate the VPs from this study into their musculoskeletal block. The model suggests several things. The students in that institution will have their own cognitive and behavioural preconditions shaped by a number of things: knowledge, their exposure to MSK problems; and their behaviours towards adopting e-learning and VPs, the behavioural precondition. The school could use this knowledge and experience to help predict how students would react to the cases. This could be done by reviewing existing attitudes towards VPs, uptake in other parts of the curriculum, and student evaluations of their e-learning programme. The directors could factor in the importance of the student experiences, in terms of 'future uptake' of the VPs.

With regard to the VPs, the institution could use the model to consider the encoded activity. The institution may wish to modify or adopt this, depending on how it fits with the local, situational, cultural and medical context as to how MSK cases present and are managed, noting the original purpose of the VPs. If the school wanted to cover particular learning objectives, these could be considered within the context of the existing encoded activity.

For the constructed activity, the institution would need to consider what the cases were used for and why, for example summative assessment, formative assessment. This would perhaps influence the environment in which students were encouraged to sit the cases. Logistical factors such as the number of computers in a teaching centre could also play a role here. The organisational approach to e-learning delivery, and assessment would help to anticipate and manage student expectation of the VPs. For example if students would not expect an MSK case in an OSCE examination, and have no summative assessment component in MSK, this will help plan the resource delivery.

The student-VP interaction is individual to an individual student. The school must acknowledge this, and perhaps if students do not wish to complete cases, or feel strongly that they are not helpful, the school could plan other activities to help manage these factors.

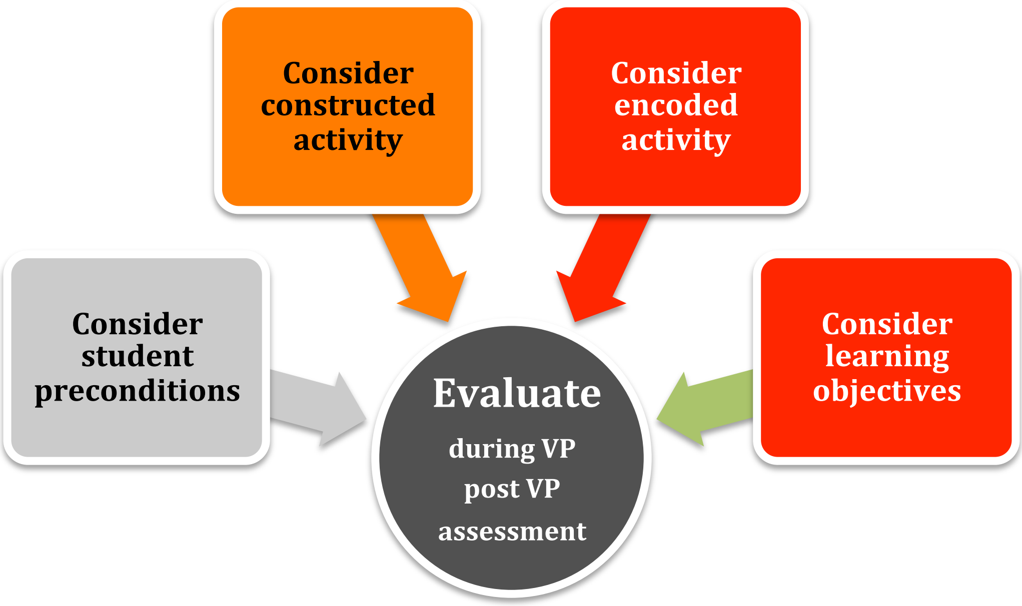
4.4. Discussion

I have described an original GT research study in medical students completing VPs that has resulted in the production of a theoretical model. The model I present is original (Bateman et al., 2013), and addresses the original research question. I will discuss the evidence supporting the practical and theoretical relevance to academics, researchers, and individuals or institutions planning to deliver VP cases. I will also discuss the limitations of research methods, within the context of existing quality markers for GT research.

At the most basic level this model explains how and why different design principles in VPs influence a student learning experience. I do not present a simple “how to” guide, but the model, grounded in the data, emphasises design as one factor in VP effectiveness. I highlight the importance of the students, encoded activity (what an author puts into a VP) and constructed activity (how an institution uses a VP). I have described six new phenomena that help provide a deeper understanding of how students interact from VPs (Table 8, p.103). An educator can review the encoded and constructed activity of his or her own cases to help plan VP development. It is possible the researchers may already be applying practices that reflect some of the elements of the model. The model does not suggest critical design variables, but requires judgement, (see Figure 16a). It is possible to abstract desirable design properties from the FG research, (see Figure 16b). I present a list of ten important design principles that emerged consistently as being important during the coding process. Within the context of the model, this research therefore presents to authors practical, straightforward advice that assists authoring of VPs.

Figure 16 a.) Using the model to help plan teaching with VPs. (b) ten examples of encoded activity

(a)



(b) Ten examples of encoded activity that can be used for VP development.

1. **Consider how e-elements are used: e.**g. limit 'Scrolling' in nodes, text size, and text amount per node. Use clear navigation strategies and signposting through the cases.
2. **Be aware of complexity and error:** this fractures the students learning experience, so use branching strategies, and complex design features with caution. Branching pathways did not emerge as particularly strong features during the case authoring.
3. **Limit VP characters.** Be aware that character numbers in the case may cause comprehension problems, in our case the consensus was three characters.
4. **Use pictures repeatedly to identify characters.** Repetitive pictures of characters helped students construct mental representations of the cases.
5. **Adopt Key Feature Problems.** The use of question materials validated to measure clinical reasoning was felt to be authentic and helpful. This is also the case for Bayesian reasoning questions.
6. **Be aware of the impact of displaying a running score.** Use of scoring systems and visible score can have positive and negative effects in different students, if possible give students a choice.
7. **Embrace the authentic presentation of normality.** This includes presenting realistic blood results, and presenting the results of 'normal' investigations authentically for interpretation.
8. **Use authentic feedback, although it may not be recognised.** Students do see feedback from VP characters as authentic, powerful, and memorable. Our students did not describe this instruction as 'feedback'.
9. **Reassure students at the outset.** Students need to understand the purpose of the VPs, and that they may feel uncomfortable. They should be told individual decisions will not result in missed learning experiences.
10. **Integrate electronic evaluations using established items.**
    Established case evaluation forms can be done simply and at low cost.

I have introduced the concept of a trade-off in design possibilities. For example, the benefits of a realistic and authentic case may cause a series of problems. Students do not appear to recognise feedback when it occurs authentically in the case (for example being corrected by a virtual physician). We coded this as the 'invisible elephant', where feedback from clinicians debating cases with trainees was not seen explicitly as 'feedback'. This may lower the 'skipping threshold', or result in students becoming 'bogged down' in the detail.

In terms of lessons for other instructional design elements, students did make specific comments about the KFP, Key Feature Problems (Page et al., 1995), see Table 6, page 99, 'decision flux'. Bayesian Reasoning questions were helpful (see quote below).

> DA: *"My best sort of my best feature about it was the fact that it was a really good teaching tool, the boxes like well explained the teaching, so the bit about the 90% sensitivity and the 90% specificity, even though I knew what they meant I realised that actually to understand how you can work out how to get it from the population, the first time I was asked the question I had completely just guessed, when I went through the feedback and learnt how it was done, and the next question I was able to get it right and I think that went really well"*
>
> Focus group 5, year 2 student

Students reaction to KFPs in the case is helpful, it supports their use in VPs, they are one of the few validated measurements of clinical reasoning skills (Schuwirth, 2009), the skills VPs are supposed to teach. Use of the assessment elements is also interesting. In Figure 16b I point out that the use of real time scoring and assessment in the cases, whilst accepted to drive learning (Newble and Jaeger, 1983), can have negative effects for the learning experience of some students. I describe this as 'points on the board', a category of mental case fracturing.

4.5. Originality and Innovative potential of this work

This research has been recognised as an important step forward in the medical literature. A commentary was published on this research publication (Bateman et al., 2013), commissioned by Professor Kevin Eva. Titled 'Research in the use of virtual patients is moving forwards by zooming out', Edelbring (2013) debates the impact of the paper on the field.

> *"In this issue of Medical Education, Bateman et al. present a study of different designs of computerised virtual patients… By clarifying these interactions between students and VPs the study can be seen as belonging to an important phase of development in which research on the use of VPs is stepping forward… Bateman et al have rightly moved on to contribute knowledge that will serve teachers and course designers in how best to design and integrate VPs into education… This study by Bateman et al is certainly timely and should inspire more research to fulfil the urgent need for systematic knowledge on various designs of VP." (Edelbring, 2013) p.544*

This model has the potential to inform future research directions. In the phenomenon skipping (Table 8) I have shown disengagement between the student and the VP. This represents a new, measurable phenomenon in VP research (Bateman et al., 2012b), and is analogous to mind wandering (Smallwood et al., 2011). When skipping occurs, the model presented suggests it relates to a combination of student preconditions, encoded activity, and constructed activity. We have seen that it appears to occur commonly. As no benchmark as yet exists for skipping, its definition and the spectrum of skipping could be identified in further research.

This research is innovative in two major areas. Firstly I have adopted a research design that created bespoke cases, using the best available technology with the explicit purpose of investigating the educational effects of the design variables I chose. Technology including using computer assisted qualitative data analysis Nvivo (QSR-International-Ltd., 2011) to interrogate and triangulate research findings against time stamped electronic decision logs from the VP software I used (DecisionSim-LLC, 2012). This to our knowledge is unprecedented in peer reviewed qualitative and quantitative VP research. This triangulation has helped inform and support both the phenomenon I describe and the validity of the findings. Secondly in terms of the VPs, by clearly identifying and publishing open-access XML cases of this thesis, I potentially set a new standard in VP education research (Bateman et al., 2013). Recent VP studies have not published either schematics (Figure 9), case information, or made research cases accessible to the healthcare professionals *or* to subscribers in journals publishing the research (Taylor et al., 2011, Edelbring et al., 2011, Botezatu et al., 2010a, Huwendiek et al., 2009b). None of these research papers reference the MedBiquitous XML standard, a recent development I have argued represents a paradigm shift in VP research (Bateman and Davies, 2011). Such publication opens avenues for scrutiny, criticism, and open peer review.

4.5.1. Research Rigour

To maintain rigour I have adhered to four 'quality markers' in GT (Bryant and Charmaz, 2007): credibility, originality, resonance and utility. This includes the use of an externally peer reviewed research protocol which has been externally scrutinised by a dedicated Education Research Committee. This work uses all the hallmarks of GT: purposeful sampling, constant comparative analysis, conceptual memoing, open and axial coding, saturation sampling, and the use of diagrams to help explain theory (Corbin and Strauss, 2008). I hold a digital audit trail, from audio recordings on file, to transcriptions, to the CAQDAS. I have presented multiple quotations to support this research. For example I am are not aware of any peer reviewed published research in this field that (1) evaluates VPs expressly constructed to evaluate design (2) allows other researchers with appropriate software to access the research cases (3) evaluates students patterns of use inside the cases using data logs. I have already described a clear acceptance of my own preconceptions and reflexivity as described in my methodological considerations (Cresswell 2007, Corbin and Strauss 2008, Malterud, 2001, see Figure 8, p. 81).

There are numerous other quality indicators in qualitative research that include perspectives from grounded theorists (Birks and Mills, 2011), critical appraisal of the qualitative research (Malterud, 2001, Kuper et al., 2008a), and its correlation with the school of GT I study (Corbin and Strauss, 20008). I argue this research has credibility as it received research funding, and has resulted in peer-reviewed publications (see **Error! Reference source not found.**) and oral presentations at national and international conferences. The open access publication makes the work accessible, and available to all authors and researchers. To this extent this work therefore addresses credibility, originality, resonance with researchers, and utility suggested to evaluate GT studies (Charmaz, 2006). I have not validated the model that I have produced. This project was not designed to validate the model presented here, but was a hypothesis generating process to build educational theory supporting the development of VPs.

4.5.2. Comparison with existing research

This model builds on existing research on VP design principles (Cook et al., 2010, Berman et al., 2009), practical authoring advice (Posel et al., 2009), curricular integration (Berman et al., 2009), theoretical principles behind VPs (Ellaway and Davies, 2011), the role of the environment (Edelbring et al., 2011), and cultural challenges for sharing cases between institutions (Fors et al., 2009). The findings support, and add to, ten general authoring recommendations produced from a thematic analysis of VPs in focus groups (Botezatu et al., 2010a), and provides a framework within which any recommendation (for example making a case authentic) can be considered within our model.

In terms of the medical education literature, this model does broadly fit with established models of e-learning practice discussed earlier (see Section 1.4, ‘VPs in the e-learning and wider educational literature’, p.51). Sara Kim and Colleagues (2006) reviewed 100 research papers, of which 40 were from the medical field which identified five core attributes of case based learning: relevant; realistic; engaging;

challenging, and instructional. There were 17 individual components, some of which fit with the findings of our research such as the 'level of the learner'. Our model embraces collaborative, shared resources accepting that a VP may have been developed by a different professional in a different context, but can be used and modified. This builds on the traditional model described by Kim et al. (2006) which described literature in which case based learning was typically used to produce bespoke resources for a single institution, which were not shared. They may not be applicable to case based learning, for example Kim and colleagues (2006) reference a paper relating to teaching veterinary histology (Eurell et al., 1999). Kim and colleagues also suggest branching is desirable, however the evidence cited for this is weak: there is no reference made to branching in the paper Kim and colleagues cite (Lechner et al., 2001). Our descriptions of branching cases in VP design draw on the work in VP typologies by Huwendiek (2009a) and describe the uncertainties that may be associated with free movement through a case. Rather than suggest branching is good, or bad, it suggests that like everything else in this model the author needs to consider the reasons for using it. Kim and colleagues also commented on the problems of evaluating case based learning, stating: "*None of the studies we reviewed used a validated measure, such as the California Critical Thinking Skills Test"*. Even the CCTST itself has problems with reliability and has been suggested to have limited validity in nursing students (Bondy et al., 2001), with again limited construct validity existing for similar surveys (Huhn et al., 2011).

Salas and colleagues (2005) provides a list of desirable factors relating to simulation based training many of which are applicable to the model, for example:

understanding training needs (student preconditions), instructional features including performance measurement (encoded activity), embedding feedback in the simulation (encoded activity : pedagogic properties). The model also highlights the relationships between the encoded and constructed activity. This reflects the literature from Salas et al. (2005) that there must be collaboration and cooperation between subject matter experts, educationalists and the institution, for simulation based technologies to be helpful. This work may be used by experts in instructional design to further refine VP development and promote further research that has been advocated in VPs (Posel et al., 2009), and successfully applied to some basic science education (Shachak et al., 2005).

Students consistently commented on the importance of feedback of reinforcing that they were not 'getting lost' inside the case, and reducing emotional discomfort. These findings are consistent with the medical education literature supporting that feedback as being the single most important feature in the success of high fidelity simulation (Issenberg et al., 2005, Salas et al., 2005).

4.5.3. Virtual Patients as vehicles for future research

This research into how students interact with VPs may be relevant to other research fields, such as those investigating mind wandering and the hidden curriculum. I described mind wandering in Section 1.5.2 (p.55), and how descriptions of skipping are analogous with task-unrelated thoughts (Smallwood et al., 2011, Bateman et al., 2012b). Our observations are in keeping with expert opinion on how these features may impact on how practicing physicians think, and make errors. Sources of error and cognitive bias are topical research areas. Students did consider the cases to be a new and novel teaching approach that would drive further learning and development, and future uptake of cases. The fact the VPs were seen as partly assessment cases may lead to pre-VP learning effects (Cilliers et al., 2012) as students prepare for completing a comprehensive assessment of topic knowledge.

The descriptions from our students indicate that one of the strengths of VPs may be their ability to address the 'hidden curriculum' a set of professional attributes described by as (Hafferty, 1998) as follows:

> *"Not all of what is taught during medical training is captured in course catalogues, class syllabi…. Indeed, a great deal of what is taught- and most of what is learned- in medical school takes place not within formal course offerings but within medicine's 'hidden curriculum'." (Hafferty, 1998)*

Our students reported features consistent with nine hidden curriculum themes from a thematic analysis of US medical undergraduates (Gaufberg et al., 2010) in particular a power hierarchy, patient dehumanisation. For example Gaufberg and colleagues identified a power hierarchy as important. So did our students as evidenced by the following quotation.

HD. *"Mr Foster [the consultant] didn't really tell you what he would have done, it was the GP that said, oh yeah that would be what I would have done."*
CB. *"Consultants don't give out praise."*
FG2, year four students

The power hierarchy seemed to have some gender effects, with students seeing the image of a more elderly male consultant daunting, and this engendered certain emotional responses such as fear (student VP interaction).

AP: "*The second case was a bit more discomforting…. And that was slightly enhanced by the fact that the face for the second one, was a man, silhouetted against the sky, intimidating… whereas…one was female, maybe a little more friendly, so the overall effect of it was a little more scary"*
FG5, Year 2 students

The patient de-humanisation described by Gaufberg and colleagues (2010) was exhibited by students (e-relationships, see quote in Table 8, page 103), as did a perceived hidden assessment of performance in their interactions with VPs (see hidden agenda, Table 8, page 103). I suggest that elements of the hidden curriculum may explain why some students link particular emotions with being corrected, for example embarrassment at being corrected by the senior team member, even in an electronic case.

Potential future directions for qualitative research in VPs do include a comparison of our work with existing theories of experiential learning (Kolb, 1984). In this seminal work, concrete experience is followed by reflective observation, abstract conceptualisation, and active experimentation. Learning can only occur through realistic experiences and reflection. Fowler (2008) describes three processes by which learning can be facilitated: a deliberate intervention by the teacher, a deliberate action by the student, or the random involvement of a third party. The theory of experiential learning suggests the concrete experiences of the students must be authentic. The methods for promoting the best forms of reflection are yet to be established in VPs. My model appears to identify barriers to learning, which include organisational factors, and pedagogic factors. For example when students feel that they do not need to know something, promotion of reflection may help. I suggest that the already established evidence supporting the integration of VPs (Berman et al., 2009) may in part relate to resource allocation for VPs, allowing VPs to experience the cases authentically. I conclude this section by giving an example of two students describing the pressures and time stresses felt from scores and timings used in the cases, with a description of moving on without reflection on the decisions made.

CB: "*I think if it was done in your own time, so you didn't feel kind of time pressured to keep moving on, not that it was too short a time or anything like that, if you can omit it you will omit it and just move on for the next thing… …if It didn't feel kind of like of a test, and it wasn't being judged against other people, maybe I might have taken more time to use it as a learning tool, and think what could I use here, and what could I not use here.*"

AL: "*So I have a different experience from you, again because I didn't look at the score.*"

Focus group 2, Year four students.

This model can potentially investigate these factors in more detail. Research from the psychology literature may also be helpful in understanding planning and applying this model. The emotions described by students and subsequent VP interactions could be considered within the framework of Vygotsky's 'Zone of Proximal Development' (ZPD) that has been applied to medical simulation technologies (Kneebone et al., 2004). The ZPD categorises tasks that can be performed unaided, tasks that can be performed with guidance (ZPD), and tasks students cannot perform.

4.5.4. Limitations

The limitations in this research relate to the cases used, the GT methodology, and the study design.

There were some design features that I did not study. These include the use of natural language text entry, the use of video, and audio in the cases. Natural language entry has been used by some researchers (Huwendiek et al., 2009b), but this is not supported by current open standards (MedBiquitous Virtual Patient Working Group, 2010) or by commercial and open source VP software (DecisionSim-LLC, 2012, Begg, 2010). The software I used may present some effects that relate to the interface. We did not use some approaches that have been adopted to teach clinical reasoning such as the 'think aloud technique' (Lee and Ryan-Wenger, 1997, Banning, 2008). This approach uses principles of metacognition, where students are encouraged to consciously verbalise thoughts as they progress through a case. Metacognition is essential to the more contemporary models of clinical reasoning

(Croskerry, 2009a), which divide automatic from conscious clinical reasoning approaches, each with its own pitfalls. Students worked through these cases as individuals. I did not explore VPs used to teach groups (Posel et al., 2009) and there were no other elearning resources included alongside the VPs that would be supported by general elearning practice (Jonassen et al., 2005). VPs are intended to promote clinical reasoning skills most commonly for individual decision making; whilst potentially a limitation, this has been suggested as a focus to improve undergraduate preparation for postgraduate working experiences (Connick et al., 2009). My cases are intended to replicate realistic clinical practice, but I did encourage students to complete the VPs using a 'closed book' assessment approach, that is not encouraging use of other resources. However, I did not enforce this. Access to additional resources in itself has been shown to influence how students utilise these sorts of teaching resources (Heijne-Penninga et al., 2008).

This research has been conducted in one institution in the UK. The population of this institution, graduate entry medical students, may in itself be an important factor. Emerging research suggests these students do not overall differ in performance with respect to age gender or academic background (Finucane et al., 2013). The fact that our students were the subjects of an intervention raises the possibility of observer effects (Holden, 2001). Our 100% completion rate for the research would support an observer effect. The fact our students were volunteers may itself not be representative of the wider population, and the research subject to a volunteer bias (Callahan et al., 2007). The findings and model do not present a simple 'how to' guide, nor do they represent a gold standard for VP research in the future. The original call in the literature from Cook and colleagues (2009) was

> *"Virtual patients are likely to play an increasing role in medical education in coming years. However, their effective use requires evidence to guide design and integration. This evidence base is currently virtually non-existent… We call upon these experts to transform this".*

This qualitative research does not transform the field, but perhaps describes the context of the field. Rather than providing a map, the model provides a compass to navigate VP design and delivery.

4.5.5. The limitations related to GT

Several schools of GT exist, and depending on the school adopted (Watling and Lingard, 2012), methodological criticisms to any application of other schools can be applied. I state my original epistemological stance, however this is controversial in GT. Much of the criticism of Corbin's GT reflects the positivist origins in the late 1960's. This is well described and debated by leading texts on qualitative and GT research theory (Denzin and Lincoln, 2011, Bryant and Charmaz, 2007). As different schools of GT exist, Charmaz (2007) herself levels criticism of Corbin for avoiding a stated epistemological stance. Juliet Corbin has explicitly avoided a prescribed 'epistemological stance' for her work on GT. To quote her (2008) book:

> *"I do not have a simple term to describe the method presented here" (Corbin and Strauss, 2008) p.7*

These challenges are faced by respected educational researchers working in the field of GT. For example Professor Glen Regehr, an editor of Advances in Health Sciences Education (in Bannister et al., 2003) and CP van der Vleuten (in Teunissen et al., 2007b) do not state an epistomolgical stance in their GT research. I share the view of Corbin that GT was developed with positivist and post-positivist roots, but does have constructivist elements. I also acknowledge that whilst perhaps the most widely used school of GT, it has been criticised. Corbin has been previously labelled as having a 'post-positivist stance' (Kennedy and Lingard, 2006), as displaying 'hints of constructivism (Watling and Lingard, 2012), but herself states:

> "*I realise there is no one "reality" out there waiting to be discovered… I agree with the constructivist viewpoint that concepts and theories are constructed by researchers." (Corbin and Strauss, 2008) p.20*

These arguments seem likely to continue in the literature.

In terms of the analysis, the emergent core phenomenon, "learning from the VP" cannot be described as ground breaking or exciting. However, this central phenomenon was not predetermined. For example, equally the focus of this research could be the 'assessment', 'professional development', 'new skills' or 'appraisal'. Any of these could form the core phenomenon. The central phenomenon 'Learning from the VP' is neither dramatic nor groundbreaking, however it has emerged from the data analysis. By making the questioning route available in the public domain for external research scrutiny, I feel this supports my coding approach

(Bateman et al. 2013). The use of multiple and potentially confusing titles such as 'e-relationships' have been subject to multiple revisions, through an iterative process. There are however many codes, and inevitably this is a potential source of confusion. The use of electronic data logs, and CAQDAS could also be criticised in a GT study as losing focus of being grounded in the research. I assert the additional data has helped us to understand and explain 'what is going on'- an approach supported by Corbin (2008):

> *"I want to emphasize that the techniques and procedures are tools, not directives. No researcher should become so obsessed with following a set of coding procedures that the fluid and dynamic nature of qualitative analysis is lost. The qualitative process should be relaxed, flexible and driven by insight gained through interaction with data." (Corbin and Strauss, 2008) p.27*

Corbin goes on to explain her own willingness to include and collect multiple forms of data, using triangulation:

> *"In any study the researcher can use any one of these alone or in combination…the desire to triangulate or obtain various types of data on the same problem… for the purposes of verifying or adding another source of data" (Corbin and Strauss, 2008) p.28*

Equally importantly, I have not presented quantitative results of student evaluations or patterns of use. These data are included to provide an alternative source of information to help understand individual student experiences (see figure 1), not to provide statistical data, which is the objective of the quantitative component described later in this thesis. Statistical analysis via aggregation of these data for quantitative analysis was not planned in our protocol, or appropriately powered. The two VPs developed use heterogeneous design variables, and students have been allocated cases using purposeful saturation sampling. This renders aggregation of these data for statistical analysis redundant.

4.6. Conclusion

This qualitative component has achieved the original aims of the study, to the extent that I have produced a model to explain the impact of VP design on student experiences, which makes a unique contribution to knowledge of VP authoring (Bateman et al. 2013). This research has been the subject of positive research commentary from other experts in the field (Edelbring, 2013), and the subject of a one page research digest summary in the journal The Clinical Teacher (2013).

I have presented a summary of the principle research findings in Figure 17. This research shows that students, authors and organisations are central to the process of VP design and delivery. Our data suggest that no single VP design exists to promote learning for all medical trainees. On the contrary, the student, and organisational factors are equally important when authoring VP cases. Whilst I acknowledge the inevitable limitations associated with this research, I argue this research informs VP design, development, and collaboration.

What was already known on this topic
• VPs are effective learning tools • There are numerous VP design typologies • It is not clear how VP design influences the experiences of students
What this research adds
• This GT research has produced a model to help understand how students interact with VPs • Encoded activity describes how an author develops a case, constructed activity refers to how an institution delivers a VP. • Student interactions with cases depends upon their prior experiences, encoded and constructed activity • VP design elements such as branching appear to have both positive and negative effects in different students, with no single design emerging as best by consensus • Students may engage in a number of behaviours that are likely to be detrimental to learning, that are measurable • Further quantitative research may be able to define the benefit of particular design variables in VPs

Figure 17 Principle research findings from the qualitative research

Section 5. VP authoring for Quantitative research study

The findings from the qualitative research support the notion that different VP design properties potentially have an impact on students completing VP cases, in terms of engagement, performance, and learning. The original plan for the research highlighted in Section 2 was for a quantitative research design to explore two important design variables. This section describes the authoring process to create cases for the second quantitative research component. I describe the choice of the two independent variables for the research, and how they are used to create different versions of four core musculoskeletal cases. This is supported by schematics showing the different case typologies. I also describe how I standardise other elements of case design, presenting an authoring template for the research cases.

5.1. The 2x2 quantitative design in four MSK topics

For the second phase of this research I have used a quantitative 2x2 multi-centre factorial study design to research individual VP design features. In this section I will describe the process used to author the four cases for this study. The context and methods for their delivery and use are discussed in the following section. A 2x2 factorial study design involves the choice of two independent variables. The design variables I have chosen to research are branching (present or absent) and structured clinical reasoning instruction (present or absent).

The clinical casesThis research was planned to run using different versions of four MSK VPs, each intended to be thirty minutes long, representing core MSK topics. The four topic choices have been informed from external peer review, a research study steering group. The four topics are pain in a single joint, pain in multiple joints, multisystem disease, and low back pain. The 2x2 factorial design is shown in Figure 18.

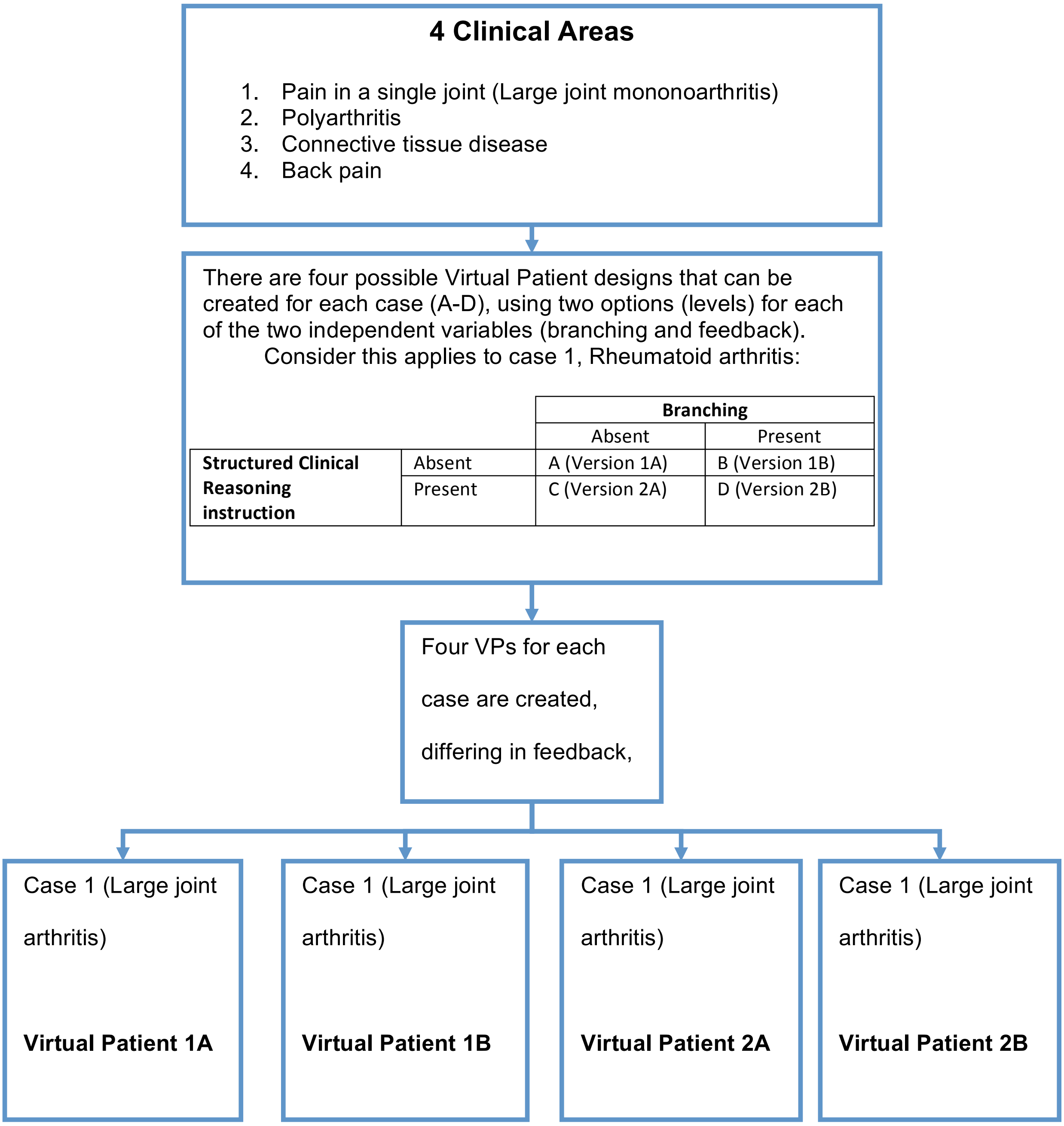


Figure 18. VP Case Design Principles: how one VP can have four different designs in the study.

5.2. Description of Branching and Structured Clinical Reasoning Instruction.

The schematics for the four possible case designs are presented later in this section.

The two independent variables are outlined in table 10.

Independent Variable **'Branching'** There are four branching points in all cases, each has three different decisions.		**Independent variable:** **'Structured clinical reasoning instruction'**	
Present	**Absent**	**Present**	**Absent**
Student makes Correct decision: Student progresses down the pathway, no difference between branching present or absent	*Student makes Correct decision:* Student progresses down the pathway, no difference between branching present or absent	Students are prompted using the SNAPPS principles (Wolpaw et al., 2009) as a series of additional nodes, but they are not given any clues to the underlying diagnosis.	Case presented in standard format.
Incorrect decision: Student goes down an alternative pathway	*Incorrect decision:* Student is taken to a step where they are given feedback that the decision was incorrect. This feedback is given authentically by a character in the case (e.g. a supervising clinician)	**S**ummarise the case, **N**arrow the differential, **A**nalysing the differential by comparing and contrasting diagnostic possibilities. **P**robe the teacher: students were encouraged to ask the teacher relevant questions	Students are left to make their own diagnostic and treatment decisions. Feedback is presented to students without them needing to ask the supervising clinician
		Students are presented with a standard reasoning problem. They are then prompted to consider Bayesian probabilities using a 'frequency grid' (Sedlmeier and Gigerenzer, 2001), see Figure 19. For example, they consider 100, or 1000 or 10000 patients to help them answer the question.	Students are presented with a standard reasoning problem.

Table 10 The two independent variables that form the four cases in a 2x2 factorial design.

An example of the SNAPPS approaches (Wolpaw et al., 2009) and frequency grid (Sedlmeier and Gigerenzer, 2001) to solve Bayesian reasoning problems is shown in Figure 19.

Figure 19 Examples of Bayes reasoning instruction to a.) help understand and interpret, b.) explain the results of Bayes reasoning problems

a.) 16 Item use of SNAPPS in the cases:

8 items to promote clinical reasoning instruction

'Based on the information to date, please try and SUMMARISE the pattern of this patients presentation and narrow the differential diagnosis'

'Try and NARROW the differential to two or three relevant possibilities.'

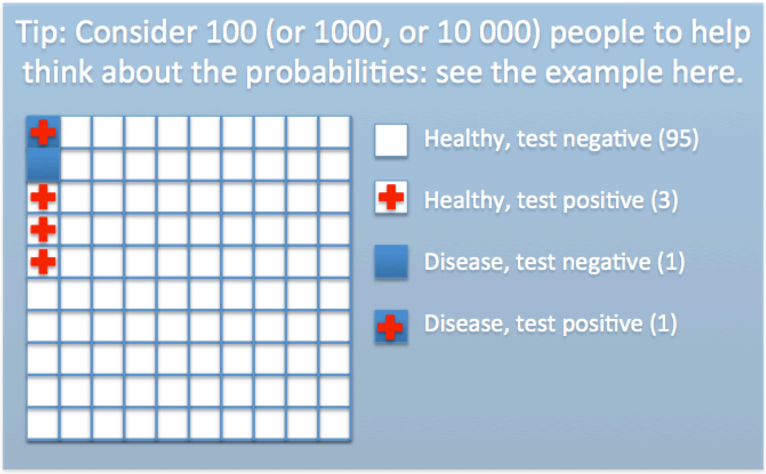
'Try to ANALYZE the differential by comparing and contrasting the possibilities'

8 items to allow students to 'probe' the supervising clinician in the case.

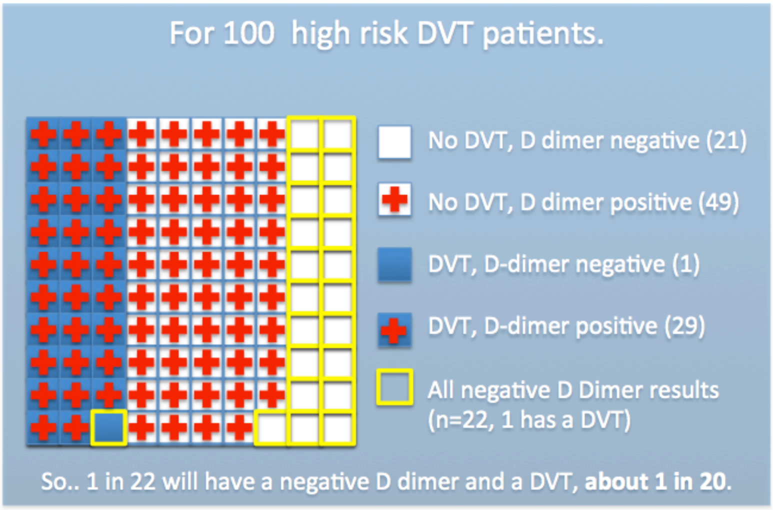
Options of questioning the supervising clinician on the rationale behind some of the decision making (this information is just presented in block format for cases where 'structured clinical reasoning instruction' is absent.

b.) Use of a frequency grid to help Bayesian Reasoning (Sedlmeier and Gigerenzer, 2001)

Prior to task: bespoke graphical representation created to promote reasoning



After the task: bespoke graphical representation to promote reasoning



5.3. The VP template.

This diagram below shows the difference between a linear (branching absent) and a branched case. It represents the nodes visited. I highlight the differences between a linear and a branching case, but also include a more detailed description of the cases (Figure 20). An example of this in practice can be seen in **Error! Reference source not found.**, p. **Error! Bookmark not defined.**.

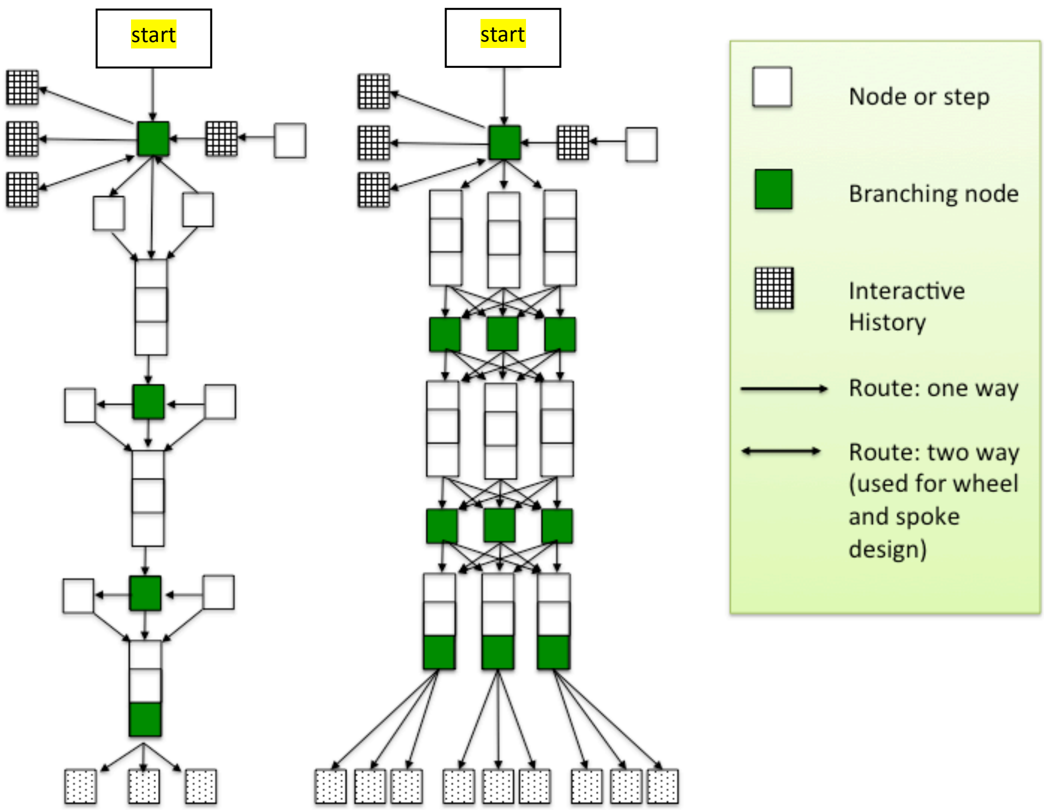
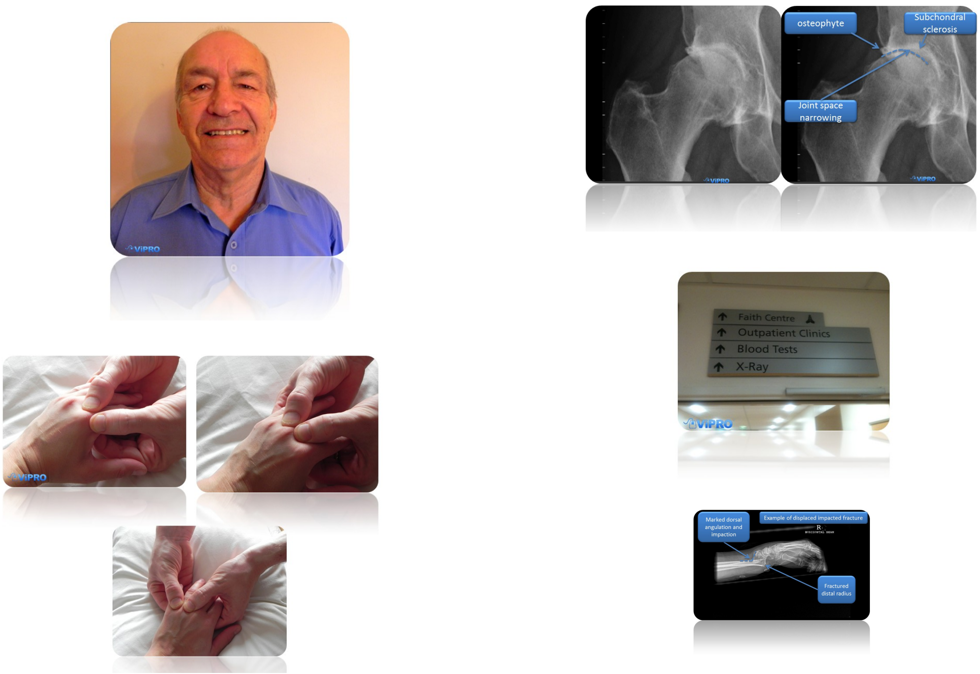


Figure 20 A linear (left) and branched (right) case schematic, showing the overall layout of the 30 minute case. The linear case has three possible endings, the branching case has nine.

5.4. Adoption of VP Media

I have used a variety of open access media in the case development, as shown below in Figure 21, each published using a Creative Commons Attribution-NonCommercial-ShareAlike licence (US National Library of Medicine and National Institute of Health, 2012).

a.) Authentic clinical and radiological images



b.) Authentic reports, letters, investigations, and settings.

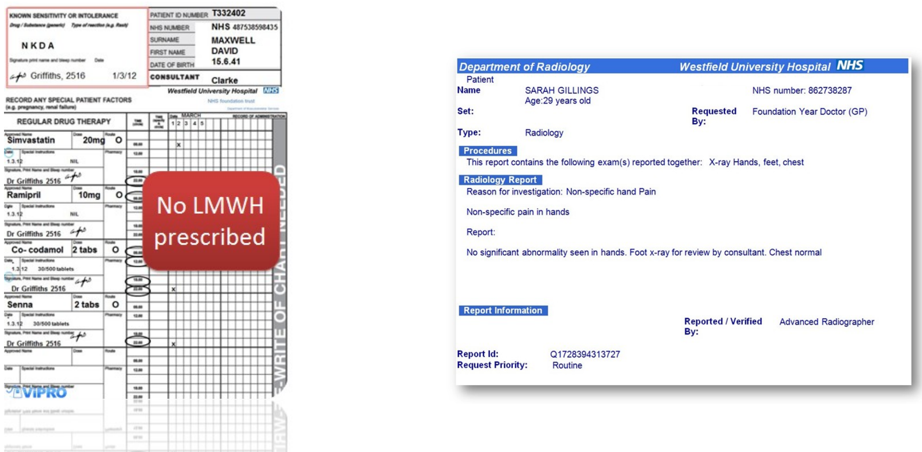


Figure 21 Examples of media used in the research process including a.)clinical and radiographic images b.) other authentic information and reports

5.5. Outcome measures: a 15 item score for each VP

As VPs are proposed to be best placed to teach clinical reasoning skills, outcomes for the study have been chosen with this in mind, and piloted in the qualitative research. I have chosen a 15 item clinical reasoning assessment utilising validated measurements of clinical reasoning skills such as Key Feature Problems (Page et al., 1995).

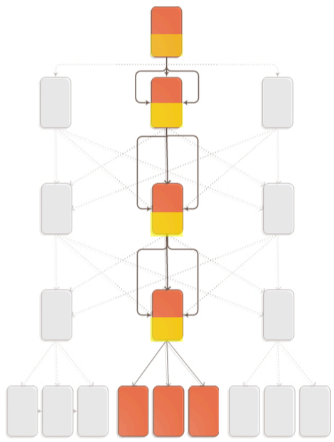
Item		*Marks awarded (where applicable)*
All Cases		
VP Question items	Key Feature Problems (x8)	8
	Clinical Decision (x4)	4
	Bayesian Statistical Question (x1)	1
	Working diagnosis (x2)	2
	Total	**15**
Other VP metrics	Time spent per case	n/a
	Time spent on questioning nodes	"
	Nodes skipped	"
	Case preference	"
	Nodes visited	"
Other metrics	Student evaluation and demographic data*	"

Table 11 Metrics collected or each VP case, inside the VP environment and following case completion

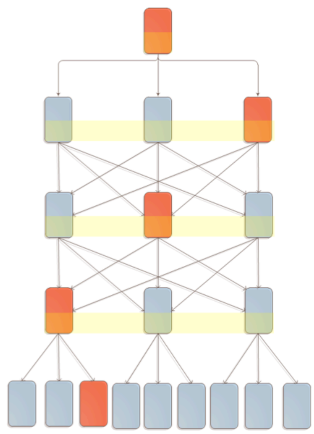
To expand on Figure 20, I present the four possible design schematics in Figure 22, illustrating how *the same* clinical reasoning questions are presented to the students at the same instances in the case.

Figure 22 The four possible case designs, 1A, 1B, 2A, 2B are all based on the same underlying template, with students receiving the same assessment questions in all four case designs

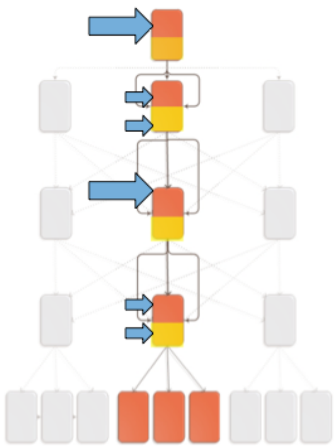
a). **Design 1A:**
Branching Absent, structured clinical reasoning instruction absent.



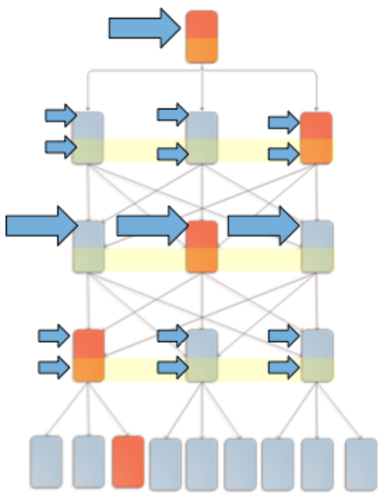
b.) **Design 1B:**
Branching present, structured clinical reasoning instruction absent



c.) **Design 2A:**
Branching absent, structured clinical reasoning instruction present



d.) **2B: Branching present,** structured clinical reasoning instruction present



Key.

Case progression*-cases are completed starting at the top working downwards as in figure 20.*
The vertical rectangles represent a sequence of steps.
Linear cases *(a,c) present a single route through the case, the branching options are greyed out, and not available to the student.*
Branching *cases are on the right (b.) and (d.) where there are four branching points with three options (4^3 routes) through the cases. To demonstrate a route through the case, the route is highlighted in orange.*
Structured clinical reasoning instruction *is represented by blue horizontal arrows. It is present in (c.) and (d.). The large blue arrows indicate a number of steps of the SNAPPS tool (Wolpaw et al., 2009), with the smaller arrows showing a single use.*
Yellow horizontal shading *represents the same in-built VP assessments used as the outcome measures (KFPs, MCQ, Bayes reasoning etc.) that occur in the same points of all four cases*

Scoring systems have been employed in the cases. Due to restrictions in the pattern of the cases, these marks were recorded across four different counters inside the VP environment, so each score (KFP, clinical decision, diagnosis) could be collected automatically.

5.6. Applying qualitative research evidence and existing theory

Both these linear and branching cases have incorporated use of instruments in the literature highlighted as being important, such as the identical use of a 'wheel and spoke' design for components of history taking. This means the user has a number of options to choose before progressing (history, past medical history, clinical examination). The VPs have been written to a structure that attempts to limit cognitive load from an individual node. This incorporates some best practice from the qualitative research. These include applying each of the ten examples described in Figure 16b, p.111 and the basic principles of multimedia design outlined in Section 1.4, VPs in the e-learning and wider educational literature, p.51.

Section 6. The Quantitative Research study

The previous chapter describes how the research study was planned, using a standardised VP design against a backdrop of a call for research into VP design properties. In Section 4 I have outlined a qualitative model that potentially predicts how students learn from VPs, but provides no quantitative data to support the impact of the design variables I studied. This section presents next step of this research process, by extending the qualitative findings into a multi-centre randomised study evaluating student experiences with VPs. In this section I describe the impact of two VP design variables student performance within a VP and their evaluation of the resource.

6.1. Problem statement and Research question.

In the earlier sections I have highlighted the issue of a lack of evidence to support how individual VP design properties influence student satisfaction and performance. The literature, published open access cases, and qualitative research have raised questions of the importance of two design components in particular. The first design variable is the use of branching in cases (present or absent). The second variable is use of evidence-based strategies to promote clinical reasoning skills based on established evidence based methods (Wolpaw et al, 2009). I describe this as 'structured clinical reasoning instruction' (present or absent).

My main research questions is:

> How do the different VP design properties of (1) branching, and (2) structured clinical reasoning instruction influence student performance and self-evaluation of VPs?

In addition I also intend to use the data to

1. Determine and explore the relationships between the results in these cases against other markers of student clinical reasoning including written and clinical summative assessment.
2. Explore how open-access cases are used by students from different institutions.
3. Explore any interaction effects between the different designs (factorial study)

6.1.1. Study Design

This is a multi-centre randomised factorial 2x2 research study evaluating the effectiveness of two independent variables on student experiences and performances using four musculoskeletal (MSK) topics, running from the 1st July 2011 to 31st December 2013. This has been conducted to an established externally peer reviewed, open access research protocol, available in **Error! Reference source not found.** (Bateman et al., 2012a). The two independent variables are: (1) branching, present or absent; (2) structured clinical reasoning instruction, present or absent (see Figure 23).

		Branching	
		Absent	Present
Structured Clinical Reasoning instruction	Absent	A (Version 1A)	B (Version 1B)
	Present	C (Version 2A)	D (Version 2B)

Figure 23 A table showing the two independent variables in a 2x2 factorial design

6.1.2. Settings and Participants

The study setting is three UK University medical schools: Warwick Medical School (WMS); The University of Birmingham Medical School (UBMS) and Keele Medical School (KMS). Due to technical challenges with timetabling of the participants sessions across three schools within the constraints of the research the setting was different at UBMS. This is shown in Table 12.

School	Students	Year group	Students invited	Course Length	Year Invited	Setting	VP delivery
WMS	GEM	~170	In individual MSK block	4 year	Year 3	University Library, sit down session	Timetabled, 4 sessions, 1 hour
UBMS	UEM and GEM	~370	By MSK block (n=6)	5 year	Year 4	Students complete in own time	No
KMS	UEM and GEM	~160	Whole year group	5 year	Year 4	University Library	Timetabled, 2 sessions, 2 hours each

Table 12 Study settings and participants (GEM=graduate entry students, UEM= undergraduate entry)

6.1.3. Virtual Patient Software

These cases were produced to be 30 minute cases using the MedBiquitous VP interoperability standard (Medbiquitous, 2010), using DecisionSim® V2.0 (DecisionSim-LLC, 2012) which uses an individual login to record and track student decisions during the cases. Cases were hosted within the University of Warwick Internet pages (go.warwick.ac.uk/msk). Evaluations where a student response was required were administered using a commercial version of the Surveymonkey (Surveymonkey Inc, Paulo Alto).

6.1.4. VP Authoring

I have outlined the four case topics and authoring strategies for the four cases in Section 5 (p.133). This includes: the case topics (Figure 18, p.134); descriptions of the two independent variables (Table 10, p.135, see also Figure 19, p.136); a template showing the case outlines of the linear and branching cases (Figure 20, p.137); examples of case media used (Figure 21, p.138); and the outcome measures integrated into the cases (Table 11, p.139). Two consultant rheumatologists, two general practitioners, and two doctors in specialist training reviewed cases, and provided written feedback on case content, style and delivery.

6.1.5. Randomisation.

I have followed the CONSORT statement on randomised trials (Moher et al., 2001). I present a flow diagram for the research protocol in Figure 24.

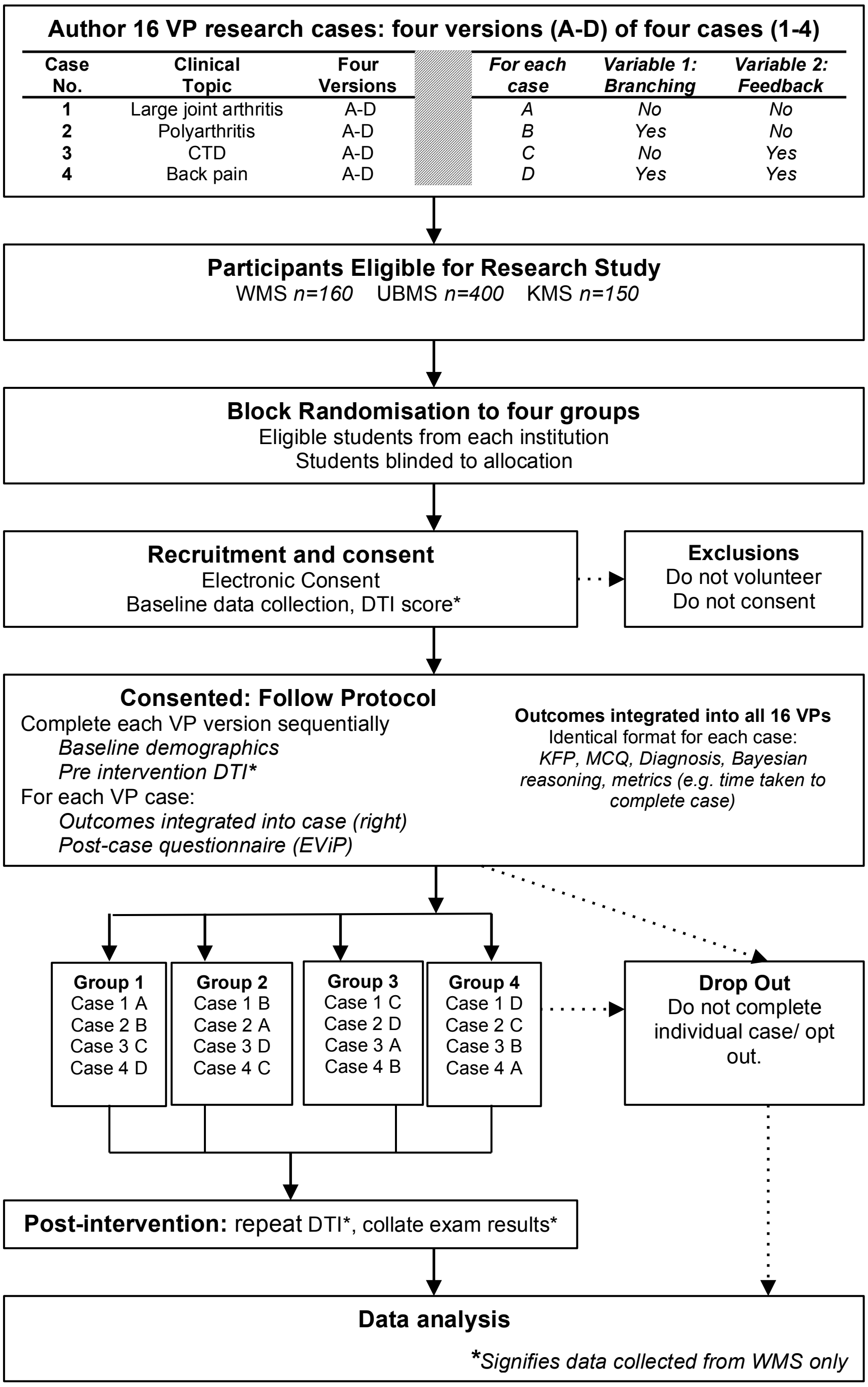


Figure 24 Flow diagram for the research protocol

I have used computerised block randomisation using random number generation to allocate students from each university cohort. Block randomisation means equal numbers of students are allocated to each group. Block randomisation is an appropriate methodology for this research as I have no other known covariates to adjust for (Kang et al., 2008), and ensures equal numbers in the four groups. The randomisation uses a computerised random number generation sequence to allocate to one of four groups. Each group can only access the cases assigned to that group. All students were randomised to the four groups prior the invitation to participate. This was done to allow students to be able to complete cases immediately after consenting to participate, without the need for a bespoke computer programmer to write an application-programming interface (API) to allocate students if they consent. The study design means that all students who complete all four cases will be exposed to all four case designs (see Figure 24, p.147). An example of which of the 16 possible VP cases a student from group 1 will be able to view is shown in Figure 25.

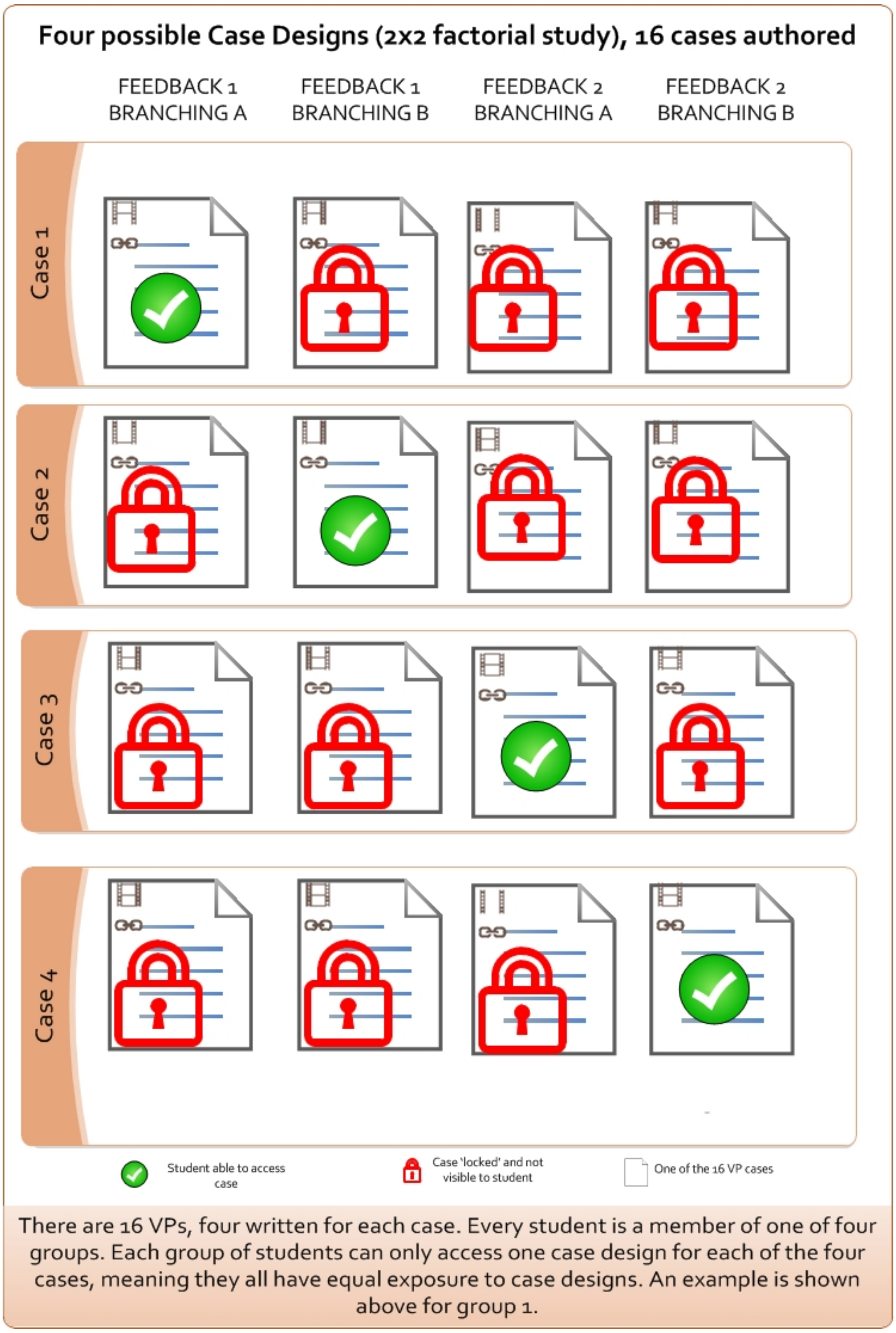


Figure 25 A schematic of the 16 case designs, showing how a student from Group 1 is exposed to all four case designs whilst working through the four cases.

6.1.6. Blinding

Students are blind to their allocation on consenting to take part in the study, and during participation. Students are not expressly told the case designs they are completing or the number of different design possibilities (four). As a researcher I was not blinded to student allocation, as this was required for facilitation of cases.

6.1.7. Timeline

A timeline for the research protocol is shown below for the three centres UBMS, KMS and WMS.

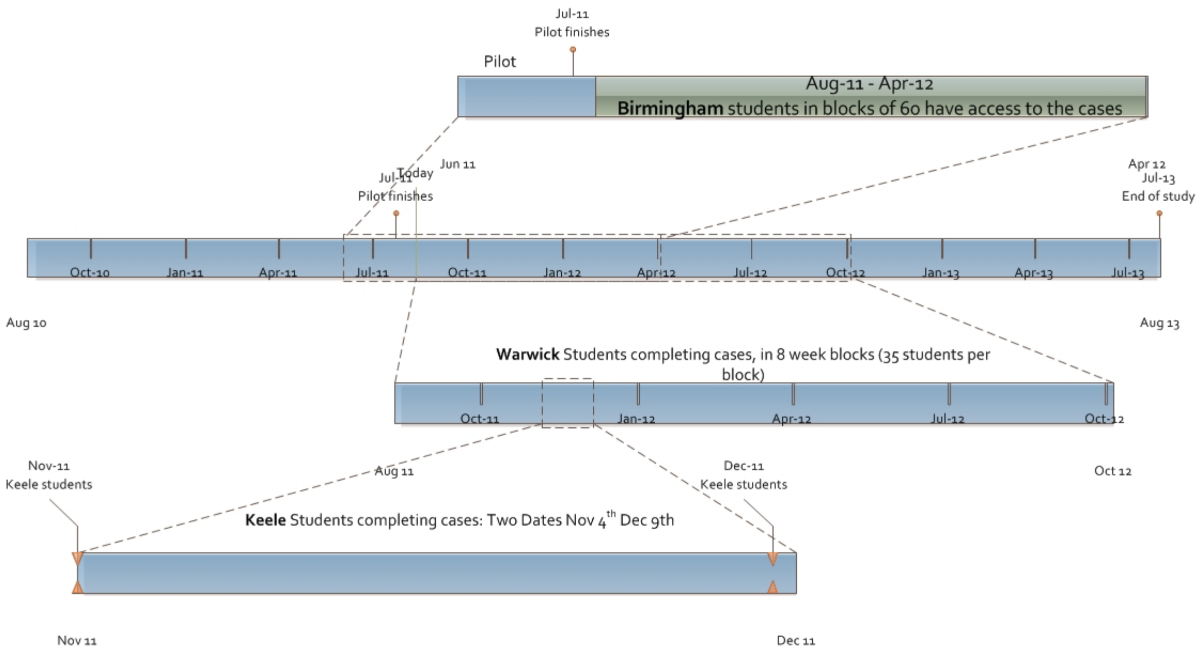


Figure 26 A schematic timeline displaying the intended introduction of VP cases in three centres in the West Midlands. Recruitment continued in Warwick until December 12, rather than October 12.

6.1.8. Recruitment and baseline data collection

I gave the same ten-minute oral presentation to all students eligible to participate in the research process. This was supported by: written participant information sheets; a web based information sheet; and an email including a participant information sheet attachment. Students who agreed to participate signed an electronic consent form. The University of Warwick Biomedical Research Ethics Committee approved the participant information sheets and consent forms. On consenting, students were considered to be participants in the study. Demographic data was collected electronically, and voluntarily from students after completing the first VP. This included gender, and graduate entry status. Students were free to withdraw at any stage.

The ‘diagnostic thinking inventory’ by Bordage et al. (1990) has been validated as a measurement of clinical reasoning against year of study. Despite its use in a pre-test post-test setting (Round, 1999), it has not been validated to measure a change in clinical reasoning skills. I considered the practical administration of such a long self reported test (completion time of 15-20 minutes) as potentially harmful to recruitment. As a result the DTI was only offered to a subset of students from one centre (WMS). Consent from the publishers (Wiley) and Professor George Bordage was given to use the metric.

6.1.9. Inclusion and Exclusion criteria

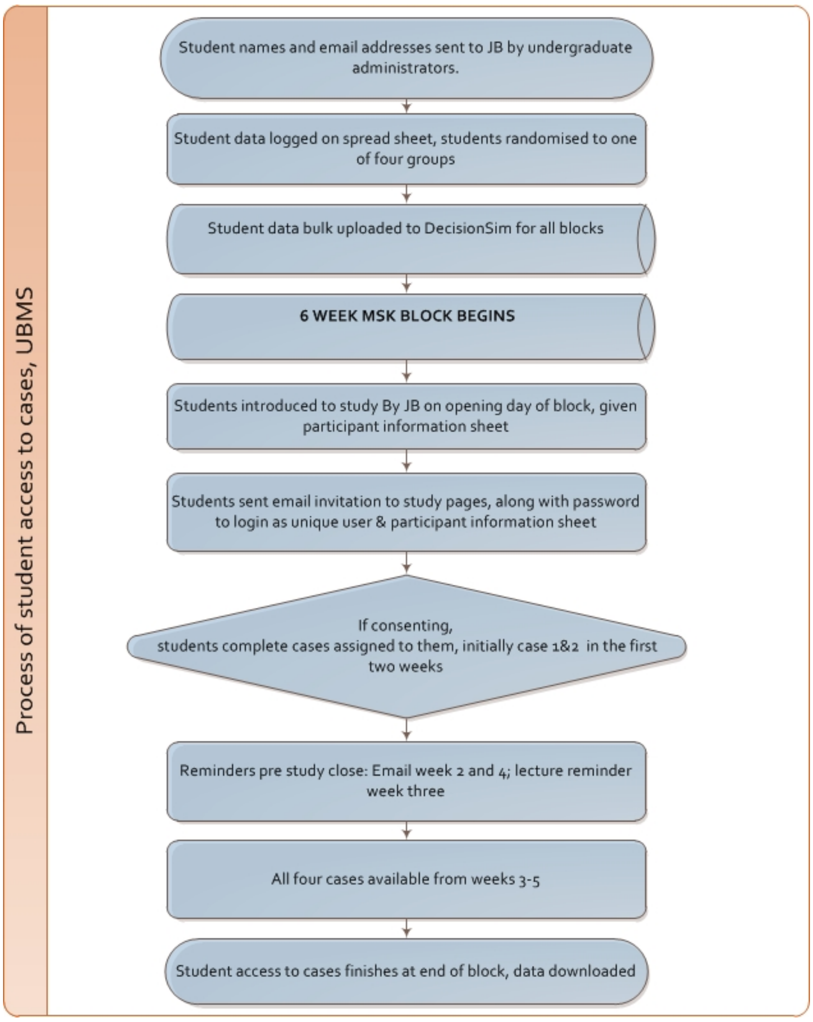
Students had to volunteer to participate, sign electronic consent forms, and be enrolled as current students on the MBChB programme. There were no specific exclusion criteria for this research, students were assumed to have the relevant information technology and language skills as part of the requirements for the MBChB programme.

6.1.10. VP delivery and data collection.

All students have a unique identifier, and had to ‘sign in’ to complete a case. This has several benefits. Firstly individual users can be tracked as they work through cases, with all of their decisions being recorded online by the software environment. This is not a function of the VP itself, but the VLE (Virtual Learning Environment) in which the VP runs (in this case DecisionSim). Secondly access to different cases can be predetermined as shown in Figure 25. The software virtual learning environment will automatically collect data in the quantitative research component for all the outcomes selected (see 6.1.11).

The delivery of the research plan was different at each centre. An example schematic for the delivery of the cases at UBMS and WMS is shown in Figure 27. At WMS the cases are delivered during four sessions during MSK block, 1 VP per session, in the hospital information technology (IT) suite. At UBMS, students complete the cases is their own time during their MSK block. Case one and two are available in weeks one and two, with all four cases available weeks three- five. At KMS, the year group accesses cases in two timetabled sessions for general medical education (two VPs per session) at the hospital IT suite. For students not able to attend two sessions the cases are delivered in a single session, in a hospital IT suite.

a.)



b.)

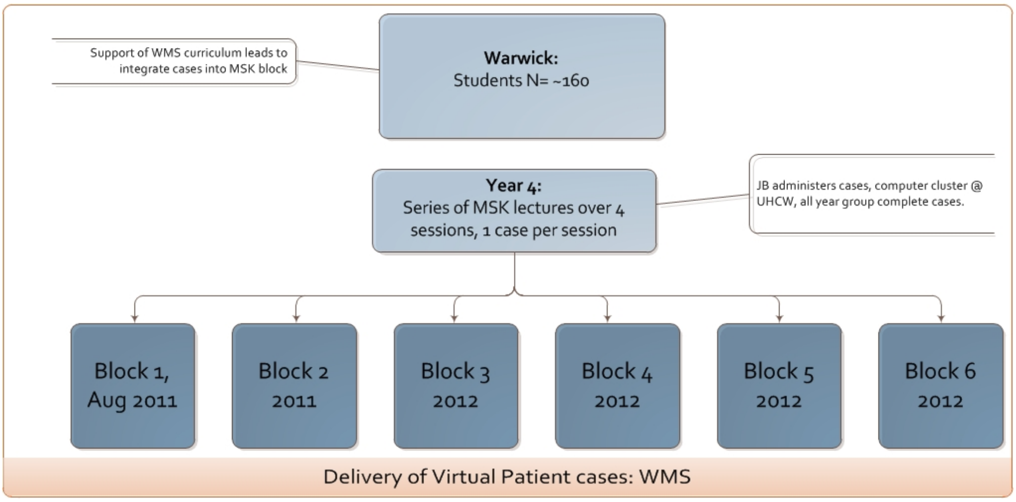


Figure 27 The delivery plan for the year groups to access the VPs at two centres (a) UBMS, and (b) WMS.

6.1.11. Outcome measures

The outcome measures are shown below in Table 13

	Outcomes for individual cases	*Timing*
Primary Outcome Measures collected for each VP		
Validated clinical reasoning assessments.	Key Feature Problems score (x/8)	
	Clinical Decision (x/4) Bayesian Statistical Question (x/1) Working diagnosis (x/2) **Total score per VP=x/15**	Mandatory: the case cannot be completed without completing these components.
Self reported evaluation (EVIP)	*EViP Questionnaire* x/55, four domains authenticity, professionalism, coaching, learning.	Optional: Students complete after each case
Case preferences	*Case Preference: reasoning (n from 4) learning (n from 4)*	Optional: On completion of all cases
	Preference of case (learning) Students selects best case for learning	Optional: On completion of all cases
	Preference of case (realism) Student selects most realistic case	Optional: On completion of all cases
Secondary Outcome measures		
Other metrics collected in electronic environment	Time spent per case (seconds) Number of nodes visited Case completion percentage. Time spent per step (seconds)	Automatically recorded during the case inside the VP learning environment
Collected from WMS Only		
Validated self reported reasoning assessment (DTI)	41 item self reported questionnaire	Immediately pre-and post- intervention
Summative assessment	IPE Written Paper: Marks out of 118, summative	
	IPE Clinical exam: summative Two Station clinical exam (64 Marks, maximum number of penalty points=48)	End of year assessment
Formative assessment	MSK written paper (Possible 51 marks), summative MSK OSCE (three stations, possible 30 marks), summative	1 week post intervention

See Bateman *et al. BMC Medical Education* 2012 **12**:62 doi:10.1186/1472-6920-12-62

Table 13 Outcome measures for the research.

6.1.12. Data analysis plan

I have collected data prospectively for the outcome variables outlined above. The statistical analysis plan was designed prior to the data collection, and was published prior to the analysis being undertaken (Bateman et al, 2012). This was set up to examine key variables for the research. This paper is included in full in **Error! Reference source not found.** (p.260).

6.1.12.1. *Original planned analysis prior to data collection*

I will present the absolute numbers of students eligible at each institution, enrolment, and case completion rates through the study. This includes presenting a participant flow diagram accounting for all students. Descriptive statistics will present the mean, standard deviation, standard error of the mean, and 95% confidence intervals for the primary and secondary analysis. The primary analysis is of performance in a 15 item clinical reasoning assessment and its components. I will present descriptive statistics of all the variables collected, along with QQ plots as visual representations of normality. Data will be presented using a standard representation of boxplots in this thesis (Tabachnick and Fidell, 2012). The box represents the 1st and 3rd quartiles (the inter-quartile range, IQR), the line in the box the median. The box 'whiskers' represent 1.5 IQR from the edge of the box. Outliers will be shown as a circle (1.5 to <3 IQR) or a star (>3 IQR).

On the presumption that the results will be normally distributed, an analysis of co-variance (ANCOVA) will be used to determine both the adjusted main-effects, and interaction effects of the two independent variables. Blocking factors in the ANCOVA will adjust for the effects of gender, institution, the case number, and graduate entry status. The ANCOVA tests will be two-sided and considered significant if p-values are <0.05. For associations between preferences with categorical rather than continuous outcomes I will use the Chi-squared test. I will explore the predictive validity of performance in the VPs in one centre, exploring correlations with summative assessment data (using Pearson's product-moment, r). The statistical analysis plan was initially planned to using the open access data analysis program 'R' (www.r-project.org), I used the statistical analysis package SPSS® version 19 (IBM, 2010) which was provided by my institution.

It is expected that as this is a voluntary study involving a significant student time commitment there will be dropout as the study progresses. The primary analysis will investigate students who complete all four VP cases. The secondary analysis will include all completed cases, irrespective of the number of cases.

6.1.12.2. *Sample Size and Errors*

I assigned a 5% difference in validated assessments of clinical reasoning skills as being an important educational effect. This research is potentially prone to type I error (incorrect rejection of the null hypothesis) and type II error (a failure to reject the null hypothesis that is false). For the power calculations I have chosen a type I error (α) of 0.05 or 5%. This means the probability of reporting a difference between the variables where none exists is 5%. For the type II error where the findings fail to reject the null hypothesis when a true difference exists is denoted 'β', which for this is 0.2, giving a power of 0.8 or 80% (power is 1- β, here I have used a power of 80%).

No gold standard clinical reasoning assessment tool exists, therefore sample size calculation is based on performance on existing KFP scores. A previous study of medical students has shown a normal distribution with a standard deviation of 1.32 when completing 15 KFPS. If we consider a student completing this number, a 5% difference in the scores would be 0.8, equivalent to a standardised effect size of approximately 0.6 (moderate to large). Using these figures the study requires 88 students to detect this difference with 80% power at the (two-sided) 5% level. If the assumption that the effect size is the same for both independent variables, 88 students would be sufficient power to detect an interaction effect between the variables that was twice as large as the main intervention effect. If the interaction between branching and structured clinical reasoning is of the same order of magnitude, a 2x2 study would require a fourfold increase in the sample size (Montgomery et al., 2003), or 352 students. For sample size calculations for student self-reported VP evaluations, a previous study has reported a standard deviation of 0.93 (Berman et al., 2009). Here a 10% difference in the evaluation scores is 0.5 on a five-point scale, equivalent to a standardised effect size of 0.5. The study needs a sample of 112 students to detect this with 80% power at the (two-sided) 5% level. As previously outlined, to detect any interaction effects between two independent variables of the same magnitude would require 448 completed VP evaluations. Both of these sample size calculations for the main effect sizes are shown in Table 14. The anticipated recruitment numbers at the outset were from a pool of over 600 potential students. Recruitment above the target given provides increasing power to detect main effects and interaction effects.

Key Feature problems					**Student self reported Evaluation scores**				
		Branching		*Total*			Branching		Total
		No	Yes				No	Yes	
Feedback	No	22	22	*44*	Feedback	No	28	28	*56*
	Yes	22	22	*44*		Yes	28	28	*56*
Total		*44*	*44*	***88***	*Total*		56	*56*	***112***

From: Bateman *et al. BMC Medical Education* 2012 **12**:62 doi:10.1186/1472-6920-12-62, See Appendix 8b

Table 14 Sample size calculation for 'key feature problems', and student self-reported evaluation scores

6.1.13. Ethical Approval

Ethical approval was granted by the Warwick Medical School Biomedical Research Ethics Committee, and was given written approval as educational research by the NHS Midlands Research Ethics Committee (see Appendix 1, ethics approval).

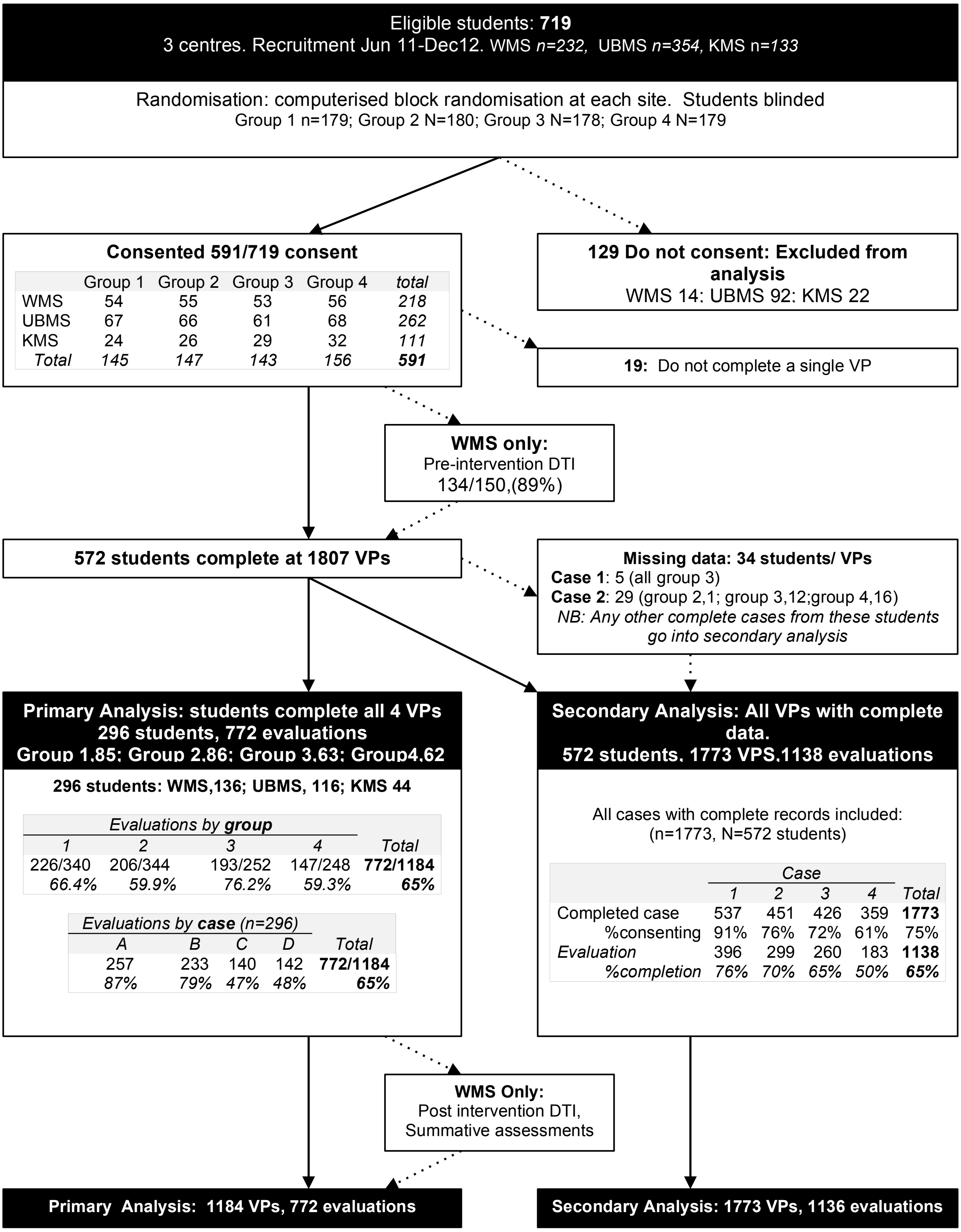
6.2. Results

In this section I describe the results of the data analysis of the study beginning with an overview of the population.

6.2.1. Participant flow diagram and overview.

A participant flow diagram is shown in line with the CONSORT (consolidated standards of reporting trials) statement (Bian and Shang, 2011). This is shown in Figure 28. From the 719 eligible students, 591 participated

Figure 28 The study participant flow through the research process reported in line with the CONSORT statement



6.2.2. Study Population

In this section I describe the demographics of the student population that was invited to, and participated in this research. A summary of the study population is presented in Table 15, p.162. In total 719 students from the three institutions were eligible to participate, of which 591 consented (82.5%). From the 719 students eligible, more were female (60.3%). Of the 591 students that consented to participate, 572 completed at least one VP case, with 296 (50.1%) completing all four cases. Volunteer and consent rates were highest at WMS (94.0%) followed by KMS (83.5%) and UBMS (74%). Overall students completed 1773 cases (mean 3.0 VPs per student), spending on average 28.6 minutes per case, returning 1229 complete evaluations (69.1% completion rate).

Eligibility for study	**Total**	**719**		
	WMS	*232*		
	UBMS	*354*		
	KMS	*133*		
	Gender			
	Male	285 (39.6%)		
	Female	434 (60.4%)		
Consent	**All students**	**591/719 (82.5%)**		
	Consent by gender	238 male (83.8% consented)		
		353 female (81.7%)		
	Consent by institution			
	WMS	218/232 (94.0%)		
	UMBS	262/354 (74.0%)		
	KMS	111/133 (83.5%)		
	Graduate status (n=591 consenting students)	GEM 249	UEM 230	unknown 105
	WMS	GEM 218	UEM 0	unknown 0
	UBMS	GEM 22	UEM 175	unknown 63
	KMS	GEM 9	UEM 55	unknown 42
VP Case Statistics	Number of MSK case topics	4		
	Total VPs evaluated	16 (four versions of each MSK topic)		
VP Completion Rates	**All students**	**1773**		
	WMS	*725*		
	UBMS	*741*		
	KMS	*307*		
Case Metrics	**Mean Score per VP, n=1773 [marks out of 15]**	**8.55 (SD 2.12)**		
	Mean time spent per VP	28.6 min (SD 13.7)		
	Mean steps/VP	56.1 (SD 7.74)		
	Mean evaluation score (/55)	44.9 (SD 5.4)		
	Mean evaluations completed /student	1.92 (SD 1.40)		
EViP Evaluations	**Total returned**	1229/1773 (69.1%)		
	Total Returned with student ID	**1138/1229** (92.6%); *or* **1138/1773 (65.2%)**		
	WMS	*478/725 (65.9% of cases)*		
	UMBS	*471/741 (63.6% of cases)*		
	KMS	*191/307 (62.2% of cases)*		
	Complete evaluations with no student ID	*91/1229 (7.4% of all evaluations)*		
Primary analysis:	**Total students completing all four cases**	**296/591 consenting students (50.1%)**		
By institution	*WMS*	136/218 (62.3%)		
	UBMS	116/262, (44.3%)		
	KMS	44/111 (39.6%)		
By randomisation group	Participation			
	Group 1	*85 (28.7%)*		
	Group 2	*86 (29.1%)*		
	Group 3	*63 (21.3%)*		
	Group 4	*62 (20.9 %)*		
	Performance (total score all four cases)	34.1/60 (all four cases), SD 5.6		
	Group 1	*34.6/60 (SD 5.1)*		
	Group 2	*33.4/60, SD 5.0)*		
	Group 3	*34.2, SD 6.3*		
	Group 4	*34.6, SD 6.4*		
Evaluations Completed	Total	772/1184 (65.2%)		
	WMS	*369/560 (65.8%)*		
	UBMS	*287/446 (64.3%)*		
	KMS	*116/176 (65.9%)*		
	Mean Evaluation scores			
	Total	45.0, SD 4.4		
	Group 1	*44.8, SD 4.4*		
	Group 2	*45.5, SD 4.3*		
	Group 3	*44.6, SD 3.8*		
	Group 4	*45.0, SD 5.1*		
Additional Data: WMS	**Summative assessment data**			
	Summative Written (IPE)	228/232 (98.3%),		
	Summative Clinical Exam	228/232 (98.3%)		
	Summative MSK OSCE	216/232 (93.1%)		
	Summative MSK Written	216/232 (93.1%)		
	Diagnostic thinking inventory			
	WMS students offered DTI	161		
	Consenting students	150		
	Pre- VP DTI; result	133/150 students (88.6%)		
	Post –VP DTI; result	85/161 students (56.7%)		

Table 15 Summary of participation data and performance of across the VP cases

6.2.3. Case completion by gender institution and randomisation

In this subsection (and Figure 29 a-h, p.165) I present data to describe and compare the study population and uptake by gender, participation, institution, completion rates, randomisation, and overall performance. There was no difference in the gender make up of the institutions (Figure 29a, p=0.156). Consent to participate differed across the three centres (P<0.001, WMS 95%, KMS 73%, UBMS 84%, Figure 29b). Each centre had different curricular integration strategies. There were no differences in case completion rate by gender (males 83.5%, females 81.3%, Figure 29c, p=0.47), or by the four possible VP case designs (N=1773, Figure 29d, p=0.795).

The primary analysis is conducted on students who completed all four cases (n=296, 1184 VPs). Students that complete all four cases are exposed to all four case designs, and therefore all groups should in principle have similar overall scores. The results confirm this, with no significant difference in overall score for students completing four cases between any of the groups (ANOVA, P= 0.613, Figure 29e). More students from groups one and two had completed all four cases (Pearson Chi-squared 29.9, p<0.001, group 1, 85; group 2, 86; group 3, 63; group 4, 62; see Figure 29f). This may reflect missing data from students in group three (n=17) and group 4 (N=17), excluding them from the final analysis. As described these exclusions did not produce any significant differences in the total scores between the groups. In the planned secondary analysis students, all completed VP cases are analysed (students completing 1-4 cases). There were no significant differences in mean VP case scores in students completing one, two, three and four VPs respectively, (ANOVA, p=0.463, Figure 29g). The four MSK case topics covered a broad range of case material in rheumatology and orthopaedics were also broadly of the same difficulty. An ANOVA that cases were not all of equal difficulty ($p<0.001$, Figure 29h). Post-hoc multiple-comparisons (Tukey) shows cases 1-3 did not differ significantly in difficulty (all $p>0.3$), with only case four having a significantly lower score than the other three cases ($p<0.001$). This difference was small with the maximum difference in mean score of 1.2/15 (Scores by case: 1,8.87; 2,8.75; 3,8.65; 4, 7.54, overall mean 8.51/15).

a.) No difference in gender split between institutions

			Gender		
			Male	Female	**Total**
University	WMS	Count	101	131	232
		% within University	43.7%	56.3%	100.0%
		% within Gender	35.2%	29.9%	32.0%
	UBMS	Count	128	226	354
		% within University	36.2%	63.8%	100.0%
		% within Gender	45.1%	52.3%	49.4%
	KMS	Count	56	77	133
		% within University	42.1%	57.9%	100.0%
		% within Gender	19.7%	17.8%	18.6%
Total		*Count*	*285*	*432*	*719*
		% within University	*39.7%*	*60.3%*	*100.0%*
		% within Gender	*100.0%*	*100.0%*	*100.0%*

Pearson Chi-Square 3.683, d.f. 2, sig 0.159

b.) Consent rates were significantly different between the three institutions, with WMS having the highest participation

			Consented	Did not consent	Total
University	WMS	Count	218	14	232
		% within University	94.0%	4.8%	100.0%
	UBMS	Count	262	92	354
		% within University	74.0%	26.0%	100.0%
	KMS	Count	111	22	133
		% within University	83.5%	16.5%	100.0%
Total		*Count*	*591*	*125*	*719*
		% within University	*82.5%*	*17.5%*	*100.0%*

Pearson Chi-Square 43.402a, df, 2, sig (2 sided) .000
a

c.) Participation by gender was equal between the two groups.

	Male	Female	Total
Consented	238 (83.5%)	353 (81.3%)	591
% within gender	*85.3%*	*81.3%*	*100%*
Did not consent	47	81	128
%within consent status	*14.7%*	*18.3%*	*100%*
Total	*285*	*434*	*719*

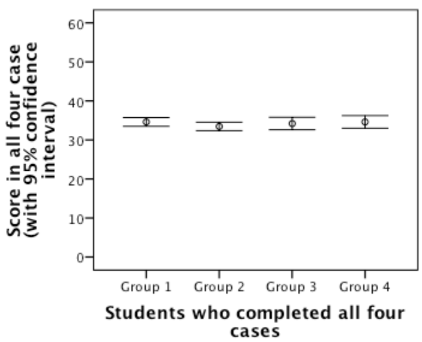
Pearson Chi- squared 0.87, 1 d.f. 1 significance [2-sided] =.471.

d.) There were no significant difference in case completion rates by case design (all completed cases, N=1773)

		Design				
		1A	1B	2A	2B	Total
Case 1	Count	136	141	117	142	536
	% within Case	25.4%	26.3%	21.8%	26.5%	100.0%
Case 2	Count	126	117	104	105	452
	% within Case	27.9%	25.9%	23.0%	23.2%	100.0%
Case 3	Count	106	107	102	111	426
	% within Case	24.9%	25.1%	23.9%	26.1%	100.0%
Case 4	Count	87	87	95	90	359
	% within Case	24.2%	24.2%	26.5%	25.1%	100.0%
Total	*Count*	*455*	*452*	*418*	*448*	*1773*
	% within case	*25.7%*	*25.5%*	*23.6%*	*25.3%*	*100.0%*

Pearson Chi squared 5.4, d.f.9, P=0.795, non significant.

e.) No difference was seen in total scores across the four groups for students that completed all four cases



ANOVA: p=0.613, no significant differences between the groups. N=296

f.) For the 296 students who completed all four cases, more students came from groups one (n=85), two (n=86), three (n=63) and four (n=62).

		Randomised Group				
		1	2	3	4	Total
Case 1	Count	85	86	63	62	296
	% within Case	*28.6%*	*29.3%*	*21.2%*	*20.9%*	*100.0%*
Case 2	Count	86	85	62	63	296
	% within Case	*29.3%*	*28.6%*	*20.9%*	*21.2%*	*100.0%*
Case 3	Count	63	62	85	86	296
	% within Case	*21.2%*	*20.9%*	*28.6%*	*29.3%*	*100.0%*
Case 4	Count	62	63	86	85	296
	% within Case	*20.9%*	*21.2%*	*29.3%*	*28.6%*	*100.0%*
Total	*Count*	*296*	*296*	*296*	*296*	*1188*

Pearson Chi squared 29.9 d.f. 9, p<0.001

g.) No significant difference in mean case performance by number of cases completed.

				95% Confidence Interval for Mean	
	N	Mean	Std. Deviation	Lower Bound	Upper Bound
1 case	79	8.71	2.136	8.23	9.19
2 cases	83	8.30	1.804	7.91	8.69
3 cases	115	8.48	1.431	8.22	8.75
4 cases	295	8.50	1.494	8.33	8.68
Total	*572*	*850*	*1.631*	*8.37*	*8.63*

ANOVA: F=0.857, Asymp. Sig. p=0.463

h.) Similar differences in case difficulty, case 4 significantly more difficult than cases 1-3, small effect size.

				95% Confidence Interval for Mean	
	N	Mean	Std. Deviation	Lower Bound	Upper Bound
Case 1	536	8.87	1.86	8.71	9.02
Case 2	452	8.75	2.07	8.56	8.94
Case 3	426	8.65	2.45	8.41	8.87
Case 4	359	7.54	1.88	7.32	7.71
Total	*1773*	*8.51*	*2.12*	*8.41*	*8.61*

ANOVA: P<0.001: Tukey post hoc tests show no differences between cases 1-3. Case 4 significantly lower score than cases 1-3, small effect size (ETA squared 0.056).

Figure 29 Comparison between the four randomised groups (a-h): a.) students invited by institution; b.) consent by institution; c.) gender participation; d.) case completion by design; e.) mean score by group; f.) groups who completed all four cases; g.) performance by cases completed; h.) case difficulty

There was no significant differences between students completing evaluations by institution: WMS, 65.9%; UBMS, 61.5%; KMS, 62.2% (Pearson's Chi-square 3.61, d.f. 2, P=0.165). Evaluation completion rates were not influenced by individual case performance (independent samples t-test, n=1767 p=0.225); gender (Pearson Chi-square 2.24, d.f.1, Asymp. 2-sided sig=0.134); graduate entry status (undergraduate entry or graduate entry, Pearson Chi-square 0.85, d.f. 1, Asymp. 2-sided sig=0.358). Students did return successively less evaluations from each case: case one 400 (74% completion); case 290 (64.0% completion), case three, 253 (59.3% completion); case four, n=177, (49.2% completion), which was significant (Pearson Chi-square 62.054, d.f. 3, Asymp. 2-sided sig.=<0.0001). This reflects number of cases completed, not case design. The student case performance did not predict if a student would complete an evaluation (independent t-test, p=0.225, see Table 16) with scores of 8.56/15 in students who completed an evaluation and 8.43/15 in those who did not.

							95% C.I. Difference	
Evaluation Completed	**N**	**Mean**	**Std. Deviation**	**Sig. (2-tailed)**	**Mean Diff.**	**S.E. Diff.**	**Lower**	**Upper**
yes	1118	8.56	2.16	0.225	0.13	0.10	-.08	0.33
no	656	8.43	2.06					

Table 16 Case scores comparing students who completed and did not complete an evaluation.

6.2.4. Descriptive statistics for case performance and evaluations

In Table 17 I present a breakdown of all 1773 completed VP cases during the study, for which there are missing data. This shows the students spent a mean 28.6 minutes per VP (SD 13.7). The overall score in the cases out of 15 marks was 8.5/15 (SD 2.1, 56.7%). For the two 'diagnosis' multiple-choice questions (MCQ), students scored 1.42 (SD 0.72, 71.0%). For the eight 'Key Feature Problems' (KFP) students scored a mean of 4.29 (SD 1.45, 53%). For the four clinical decisions (one from three), students scored 2.65/4 (SD 1.0, 66.2%). For the one Bayesian reasoning question, students scored 0.15 (15%). The overall evaluation score (EViP self-reported evaluation, 55 possible marks with 55 representing the best case) was 44.9 (SD 5.43, 81.6%). In Table 17 I also present the descriptive statics for the primary analysis, the cohort that completed all four VPs (n=296 students).

						Skewness		Kurtosis	
	N	*Minimum*	*Maximum*	*Mean*	*Std. Deviation*	Statistic	Std. Error	Statistic	Std. Error
All completed cases: N= 572 students. N=1773 cases completed									
Case Metrics									
Minutes	1773	3.25	118	28.65	13.68	1.93	.058	5.88	0.12
Steps	1773	34	88	56.10	7.744	0.020	.058	-0.20	0.12
Diagnosis	1773	0	2	1.42 (71.0%)	0.72	-0.81	.058	-0.65	0.12
KFP	1773	0	8	4.29 (53.6%)	1.46	-0.05	.058	-0.21	0.12
Clinical Decisions	1773	0	4	2.65 (66.2%)	1.00	-0.38	.058	-0.45	0.12
Bayes Reasoning	1773	0	1	0.15 (15%)	0.36	1.96	.058	1.83	0.12
Total score	1773	2	15	8.51 (56.7%)	2.13	-0.22	.058	-0.09	0.12
Evaluation Scores									
authenticity (/10)	1138	2	10	8.08 (80.1%)	1.26	-0.63	.073	1.55	0.14
professionalism (/20)	1138	4	20	16.04 (80.0%)	2.19	-0.52	.073	2.26	0.14
coaching (/15)	1138	3	15	12.54 (83.6%)	1.76	-1.33	.073	7.23	0.14
Learning (/10)	1138	2	10	8.20 (82.0%)	1.21	-1.49	.073	7.58	0.14
Total (/55)	1138	11	55	44.94 (81.6%)	5.44	-.722	.074	3.10	0.14
Students completing all four cases: N=296 students, N=1184 cases									
Case Metrics									
Minutes	1184	3.25	118	28.75	13.70	1.89	.071	5.70	.142
Steps	1184	34	88	56.70	7.89	0.11	.071	-0.21	.142
Diagnosis	1184	0	2	1.35	0.74	-0.67	.071	-0.90	.142
KFP	1184	0.33	8	4.33	1.43	-0.05	.071	-0.20	.142
Clinical Decisions	1184	0	4	2.70	0.98	-0.40	.071	-0.49	.142
Bayes Reasoning	1184	0	1	0.15	0.36	1.91	.071	1.67	.142
Total score	1184	2	15	8.53	2.11	-0.19	.071	-0.17	.142
Evaluation Scores									
authenticity (/10)	772	2	10	8.11	1.25	-0.65	.087	1.61	0.18
professionalism (/20)	772	4	20	16.06	2.17	-0.54	.087	2.74	0.18
coaching (/15)	772	3	15	12.55	1.68	-0.96	.087	4.96	0.18
Learning (/10)	772	2	10	8.24	1.11	-1.13	.087	5.89	0.18
Total (/55)	772	11	55	44.98	5.24	-0.72	.088	3.42	.18

Table 17 Descriptive statistics for all completed cases and evaluations.

The evaluation scores shown in Table 17 are the 1184 evaluations that can be associated with an individual user. For reference I have included the EViP questionnaire items in full with responses and SD in Table 18.

Question	EViP domain: introductory text presented to students	Question Text	Mean	Std. Deviation
EVIP1-Auth	'Authenticity of patient encounter and the consultation'	While working on this case, I felt I had to make the same decisions a doctor would make in real life.	4.21	.638
EVIP2-Auth		While working on this case, I felt I were the doctor caring for this patient.	3.88	.784
EVIP3-Prof	'Professional approach in the consultation'	While working through this case, I was actively engaged in gathering the information (e.g., history questions, physical exams, lab tests) I needed, to characterize the patient's problem.	4.04	.669
EVIP4-Prof		While working through this case, I was actively engaged in revising my initial image of the patient's problem as new information became available.	4.12	.670
EVIP5-Prof		While working through this case, I was actively engaged in creating a short summary of the patient's problem using medical terms.	3.67	.887
EVIP6-Prof		While working through this case, I was actively engaged in thinking about which findings supported or refuted each diagnosis in my differential diagnosis.	4.19	.623
EVIP7-Coach	'Coaching during consultation'	I felt that the case was at the appropriate level of difficulty for my level of training.	4.10	.758
EVIP8-Coach		The questions I was asked while working through this case were helpful in enhancing my diagnostic reasoning in this case.	4.21	.608
EVIP9-Coach		The feedback I received was helpful in enhancing my diagnostic reasoning in this case.	4.22	.653
EVIP10-Learn	'Learning effect of consultation'	While working through this case, I was actively engaged in creating a short summary of the patient's problem using medical terms.	4.07	.677
EVIP11-Learn		After completing this case, I feel better prepared to confirm a diagnosis and exclude differential diagnoses in a real life patient with this complaint.	4.13	.635
EVIP12-Global	'Overall judgement'	Overall, working through this case was a worthwhile learning experience*: Not used, substituted with questions relating to design.*	N/A	
Total: All returned EViP questionnaires- n=1229)			44.84 /55	5.33
Total linked to an individual student ID (n=1138)			*44.94/55*	*5.44*

Note. Although widely adopted, this questionnaire has not been formally validated. The reliability of the items is supported by a Cronbach's alpha of 0.894 (95% C.I. 0.89-0.90). This demonstrates good internal consistency, supporting the questionnaire items measure the same construct: a value of 0.7 or higher is acceptable, with values of over 0.8 preferred (Tabachnick and Fidell, 2007). Calculating Cronbach's alpha when removing any one of the 11 questions resulted in a lower Cronbach's alpha in every case, suggesting no question items should be removed. The complete inter-item correlation matrix is shown in appendix 5.

Table 18 Outline of the EViP questionnaire items (Huwendiek and de Leng, 2010), and metrics from all 1229 completed evaluations returned. Each item is a 1-5 Likert scale from strongly disagree- strongly agree.

For completeness I present the descriptive statistics for all completed cases in Table 19. In the study protocol, and participant flow diagram (Figure 28, p.160), the primary analysis is based on students completing all four VP cases. I have presented the case metrics in boxplots in Table 20, which supports the similarities between the cases for time (Table 20a), steps (Table 20b, clinical reasoning scores (Table 20c), overall case score (Table 20d) and evaluation score (Table 20e). I have used a standard representation of boxplots in this thesis, see 6.1.12.1, p.155.

						95% C.I. for Mean			
		N	Mean	Std. Deviation	Std. Error	Lower Bound	Upper Bound	Minimum	Maximum
Time (Minutes)	Case 1	537	28.32	13.29	0.56	27.21	29.42	1.20	96.00
	Case 2	451	28.81	14.06	0.63	27.56	30.05	3.68	114.00
	Case 3	426	27.54	13.04	0.62	26.33	28.75	1.18	88.00
	Case 4	359	27.54	15.12	0.77	26.03	29.06	1.52	118.00
	Total	**1773**	**28.10**	**13.83**	**0.32**	**27.48**	**28.73**	**1.18**	**118.00**
Steps	Case 1	537	55.95	7.067	.305	55.35	56.55	36	88
	Case 2	451	55.55	6.631	.312	54.93	56.16	35	73
	Case 3	426	51.49	5.977	.290	50.92	52.06	34	67
	Case 4	359	62.49	7.597	.401	61.70	63.28	37	74
	Total	**1773**	**56.10**	**7.744**	**.184**	**55.74**	**56.46**	**34**	**88**
Diagnosis	Case 1	537	1.88	.336	.015	1.85	1.90	0	2
	Case 2	451	1.57	.601	.028	1.51	1.62	0	2
	Case 3	426	1.45	.643	.031	1.39	1.51	0	2
	Case 4	359	.50	.501	.026	.45	.56	0	2
	Total	**1773**	**1.42**	**.717**	**.017**	**1.38**	**1.45**	**0**	**2**
KFP	Case 1	537	4.32	1.39	0.06	4.21	4.44	0	8
	Case 2	451	4.51	1.45	0.07	4.37	4.64	0.33	8
	Case 3	426	4.30	1.62	0.08	4.15	4.46	0.33	8
	Case 4	359	3.97	1.30	0.07	3.83	4.11	0.66	7
	Total	**1773**	**4.29**	**1.46**	**0.03**	**4.23**	**4.36**	**0**	**7**
Clinical Decisions	Case 1	537	2.64	0.98	0.04	2.55	2.72	0	4
	Case 2	451	2.51	0.92	0.04	2.42	2.59	0	4
	Case 3	426	2.63	1.13	0.06	2.52	2.73	0	4
	Case 4	359	2.87	0.91	0.05	2.78	2.97	0	4
	Total	**1773**	**2.65**	**1.00**	**0.02**	**2.60**	**2.69**	**0**	**4**
Bayes Reasoning	Case 1	537	0.03	0.18	0.01	0.02	0.05	0	1
	Case 2	451	0.17	0.38	0.02	0.14	0.21	0	1
	Case 3	426	0.26	0.44	0.02	0.22	0.30	0	1
	Case 4	359	0.17	0.37	0.02	0.13	0.21	0	1
	Total	**1773**	**0.15**	**0.36**	**0.01**	**0.13**	**0.17**	**0**	**1**
Total score	Case 1	537	8.87	1.86	0.08	8.71	9.02	2	13.66
	Case 2	451	8.75	2.07	0.10	8.56	8.94	2	14
	Case 3	426	8.65	2.43	0.12	8.41	8.88	2	15
	Case 4	359	7.52	1.88	0.10	7.32	7.71	2	12.66
	Total	**1773**	**8.51**	**2.13**	**0.05**	**8.41**	**8.61**	**2**	**15**
Evaluation (authenticity) /10	Case 1	396	8.08	1.261	.062	7.96	8.20	2	10
	Case 2	299	7.98	1.303	.075	7.83	8.13	2	10
	Case 3	260	8.10	1.299	.080	7.94	8.25	4	10
	Case 4	183	8.30	1.086	.080	8.14	8.46	5	10
	Total	1138	8.09	1.258	.037	8.02	8.16	2	10
Evaluation (professionalism) /20	Case 1	396	15.85	2.177	.107	15.64	16.06	6	20
	Case 2	299	16.01	2.119	.122	15.77	16.25	4	20
	Case 3	260	16.15	2.435	.151	15.86	16.45	5	20
	Case 4	183	16.48	1.915	.142	16.20	16.76	12	20
	Total	1138	16.06	2.193	.064	15.93	16.19	4	20
Evaluation (coaching) /15	Case 1	396	12.80	1.583	.078	12.64	12.95	4	15
	Case 2	299	12.33	2.066	.119	12.09	12.56	3	15
	Case 3	260	12.33	1.797	.111	12.11	12.55	7	15
	Case 4	183	12.66	1.463	.108	12.44	12.87	8	15
	Total	1138	12.55	1.763	.052	12.45	12.65	3	15
Evaluation (Learning) /10	Case 1	396	8.26	1.195	.059	8.14	8.37	2	10
	Case 2	299	8.11	1.311	.076	7.96	8.26	2	10
	Case 3	260	8.17	1.201	.074	8.03	8.32	4	10
	Case 4	183	8.36	1.011	.075	8.21	8.51	5	10
	Total	1138	8.22	1.203	.035	8.15	8.29	2	10
Evaluation (Total)	Case 1	396	45.06	5.296	.270	44.53	45.59	15	55
	Case 2	299	44.43	5.598	.323	43.79	45.07	11	55
	Case 3	260	44.75	5.833	.361	44.04	45.47	22	55
	Case 4	183	45.80	4.752	.351	45.10	46.49	31	55
	Total	1138	44.94	5.435	.162	44.63	45.26	11	55

Table 19 Performance metrics for the four cases sat by students.

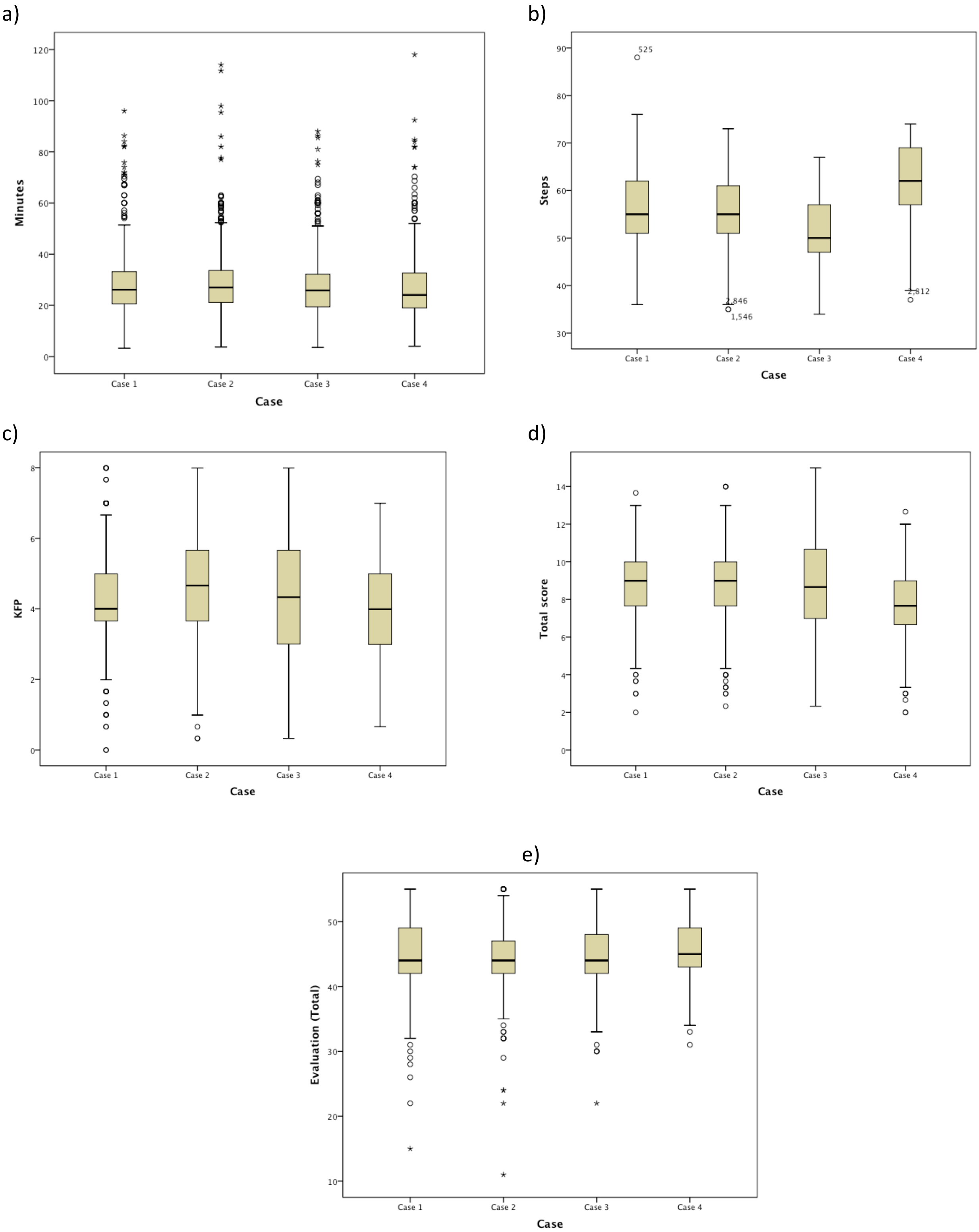


Table 20 Boxplots for the case performance metrics [2]

[2] *Footnote. A standard representation has been used Tabachnick, B. G. & Fidell, L. S. 2012. Using multivariate statistics, Boston, Mass. ; London, Pearson..The box represents the 1st and 3rd quartiles, or the inter-quartile range (IQR). The line in the box represents the median. The 'whiskers' represent' 1.5 IQR from the edge of the box, data outside these points are represented as either a circle of a star. A circle indicates an outlier further than 1.5 IQR from the edge of the box, but greater less than 3 IQR away. A star demotes an outlier equal to or greater than 3 interquartile ranges from the edge of the box.*

6.2.4.1. Missing data

In 32 cases performance data was unavailable due to unforeseen technical problems, and student performance scores not being logged in the e-learning environment. The 32 cases involved 32 different students from all centres. The source of this error was not clear, likely to be multifactorial including software, hardware and authoring issues. As a proportion of all completed cases this is small, 1.8% (32/(1773+32)). This is represented on the participant flow diagram (Figure 28, page 160). These students were excluded from the primary analysis, as they did not have complete records for all four cases, and this has contributed to some of the differences seen in group numbers. As shown in Figure 29e (p.165), there were more students in groups one and two. When accounting for missing data, there were no differences in case completion rates by group for case 2, 3 and 4 (Pearson Chi-squared p=0.19 for case 2; p=0.46 for case 3; p=0.44 for case 4; see Table 21). There were significant differences in completion rates by randomization for case 1 only, with case 2A (structured clinical reasoning present, branching absent) having fewer completed cases than the other case designs (Pearson Chi-squared p=0.008, see Table 21, p. 174). The absolute numbers for case 1 were small (142,136,122,142). Where data was missing from one case, the other cases students completed were included in the secondary analysis.

		Case 1 Completed		
		Yes	No	Total
Randomised to group	Group 1	136	9	145
	Group 2	142	5	147
	Group 3	122*	21	143
	Group 4	142	14	156
Total		542	50	591
	Pearson Chi-Square	11.9, d.f.3, p=.008		

		Case 2 Completed		
		1.00	2.00	Total
Randomised to group	Group 1	116	29	145
	Group 2	128	19	147
	Group 3	117*	26	143
	Group 4	120*	36	156
Total		481	111	591
	Pearson Chi-Square	4.6, d.f.3, p=0.19		

		Case 3 completed		
		1.00	2.00	Total
Randomised to group	Group 1	101	44	145
	Group 2	112	35	147
	Group 3	106	37	143
	Group 4	107	49	156
Total		426	166	591
	Pearson Chi-Square 2.6, d.f.3, p=0.46			

		Case 4 Completed		
		Yes	No	Total
Randomised to group	Group 1	89	56	145
	Group 2	96	52	148
	Group 3	87	56	143
	Group 4	87	69	156
Total		359	233	591
	Pearson Chi-Square	2.69, d.f.3, p=0.44		

*Includes cases where case completed but data missing (n=5 for case 1, n=29 for case 4.

Table 21 Case completion rates for each of the four cases (the four clinical topics) by randomised group.

There was no ‘incomplete’ case evaluation data as students could only submit completed cases. There was a separate (identical) evaluation for each case, however to link the evaluation with the student, the student had to type their 7-9 digit student ID. I could not match student IDs for 93/1229 evaluations (7.6%) meaning 1136 complete evaluations were usable, response rate 1136/1773 (64.0%). For students completing all four cases, 772/1184 complete evaluations were received (65.2%) students.

6.2.4.2. *Data from WMS only*

Descriptive statistics for the WMS assessment results are shown in Table 22.

This includes the diagnostic thinking inventory results (Bordage et al., 1990) pre and post intervention, and a written summative assessment paper (IPE written), a summative two station clinical exam (IPE clinical), penalty points accrued in the summative clinical exam (IPE penalty point) and the total score in a MSK assessment.

	N	Minimum	Maximum	Mean	SD
End of block assessment					
MSK written (/51)	216	34.0	51.0	45.6	3.5
MSK OSCE (/30)	216	18.0	29.0	24.3	2.3
End of Year assessment					
End of year written (/118)	228	56.0	111.0	88.7	10.8
End of year clinical (/80, negatively marked)	228	-10.0	58.0	29.9	10.3
Pre-post VP assessments*					
DTI Pre-VP	133§	105.0	208.0	163.6	17.6
DTI Post-VP	85	112.0	221.0	163.6	18.0

*Note: the DTI was offered to 161 consecutive students participating at the WMS site. 71/232 students were not offered the DTI.

Table 22 Descriptive Statistics for additional data collected from WMS.

There were no significant differences in summative clinical and written assessment scores between students who consented to participate and those that did not (all p>0.6), or with students who completed ≥2 VPs compared to those who completed 0-1 VPs (all P>0.6). This is shown in Table 23, p.176. Furthermore as with the cohort as a whole, performance and evaluation was not significantly different in those that completed a 0-1 cases and those that completed ≥2 VPs (p>0.25, independent samples t-test, see Table 23).

	Cases completed	N	Mean	Std. Deviation	Sig. (2-tailed)	Mean Difference	Lower	Upper
MSK Written*	>= 2	185	45.6	3.4	.386	-.55	-1.9	0.8
	< 2	31	46.1	3.9				
MSK OSCE*	>= 2	185	24.3	2.3	.932	-.069	-0.9	.9
	< 2	31	24.3	2.3				
IPE written (out of 118)+	>= 2	193	88.9	10.7	.824	0.60	-3.5	4.4
	< 2	35	88.3	11.7				
NEW IPE clinical+	>= 2	193	29.8	10.1	.860	0.44	-3.4	4.1
	< 2	35	29.6	11.9				
Mean VP Evaluation¶	>= 2	178	43.2	5.1	.441	1.06	-1.6	3.7
	< 2	17	42.1	8.5				
Mean VP score¶	>= 2	194	8.0	1.4	.241	0.41	-0.3	1.1
	1 or 2	21	7.6	2.1	.			

*MSK written paper, and MSK OSCE complete results, N=216/232
+IPE Written and IPE clinical complete results N=228/232
¶n=215. [215/218 consenting students consented and completed cases from WMS]

Table 23 Case completion rates did not predict performance in exams, and VP performance metrics did not predict who would go on to complete further cases.

6.2.5. Tests of Normality

I have conducted tests of normality for each of the dependent variables including patterns of use and time spent in the cases, and performance. I have used a QQ plot (or quantile-quantile plot) to display the normality of the distributions, plotting the observed values (x-axis) against a normal distribution (y-axis). QQ plots will highlight individual outliers and skewed distributions. The VP performance scores (Figure 30a, b), KFP scores (Figure 30 c, d), and evaluations (Figure 30 e, f) can be seen to be approximately normally distributed from the QQ plots and histograms (p.178). The evaluation scores are also normally distributed, and are skewed to the right (as evidenced from the histogram, and the QQ plot). This means students were more likely to report a more positive evaluation. The patterns of use metrics are also largely normally distributed. This is shown in QQ plots and histograms in Figure 31 a-f, page 179. This shows QQ plots and histograms for the number of steps taken and time spent per case. The number of steps taken during the case (Figure 31a and Figure 31b) is normally distributed. The histogram and QQ plot show the time spent per case does show a normal distribution with a skew to the left (Figure 31c, and

Figure 31d). As expected, by taking a log of the time spent, the skew is largely removed (Figure 31e and Figure 31f). The assessment data I have collected from students from one centre (WMS) is also normally distributed in terms of the IPE clinical examination score, and their musculoskeletal exam. This is shown in Figure 32, p.180.

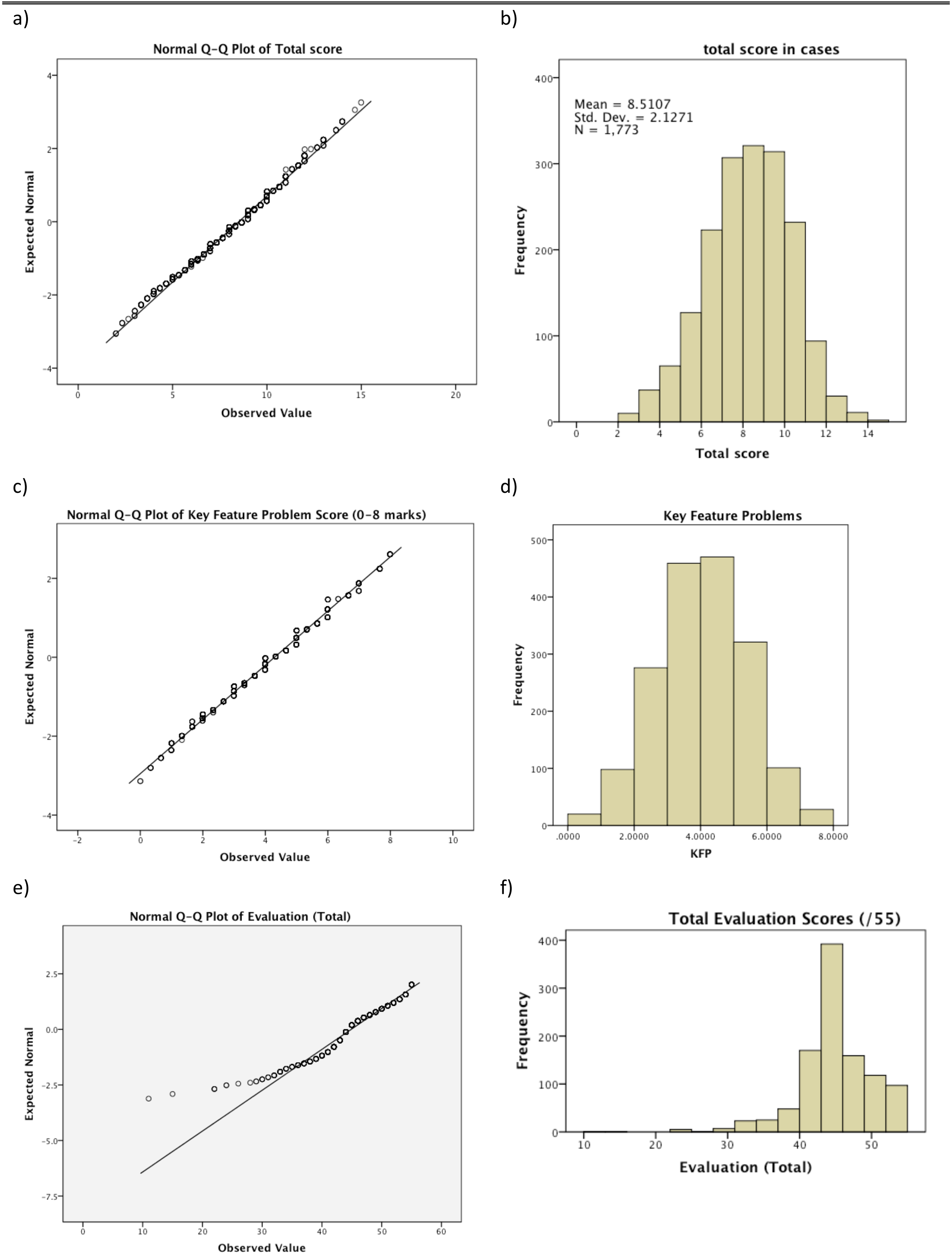


Figure 30 QQ plots and histograms for VP performance scores (a,b), KFPs (c,d), and EViP evaluations (e,f), with the suggestion of a skew to the right in the QQ pots.

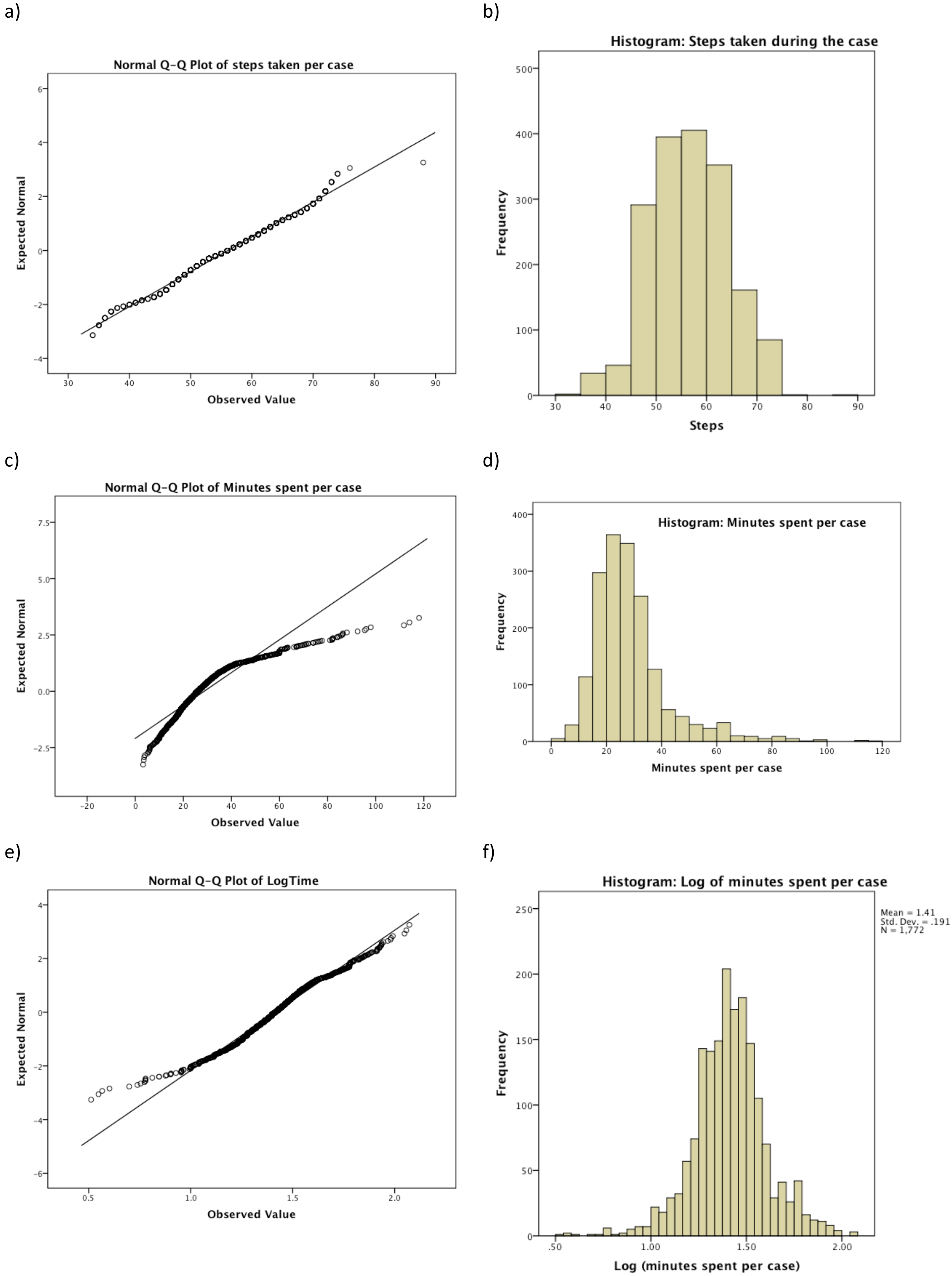


Figure 31 QQ plots and histograms showing steps taken per case (a,b) time taken (c,d), and log(time) (e,f) The distribution of minutes per case suggests a long tail at each side of the distribution.

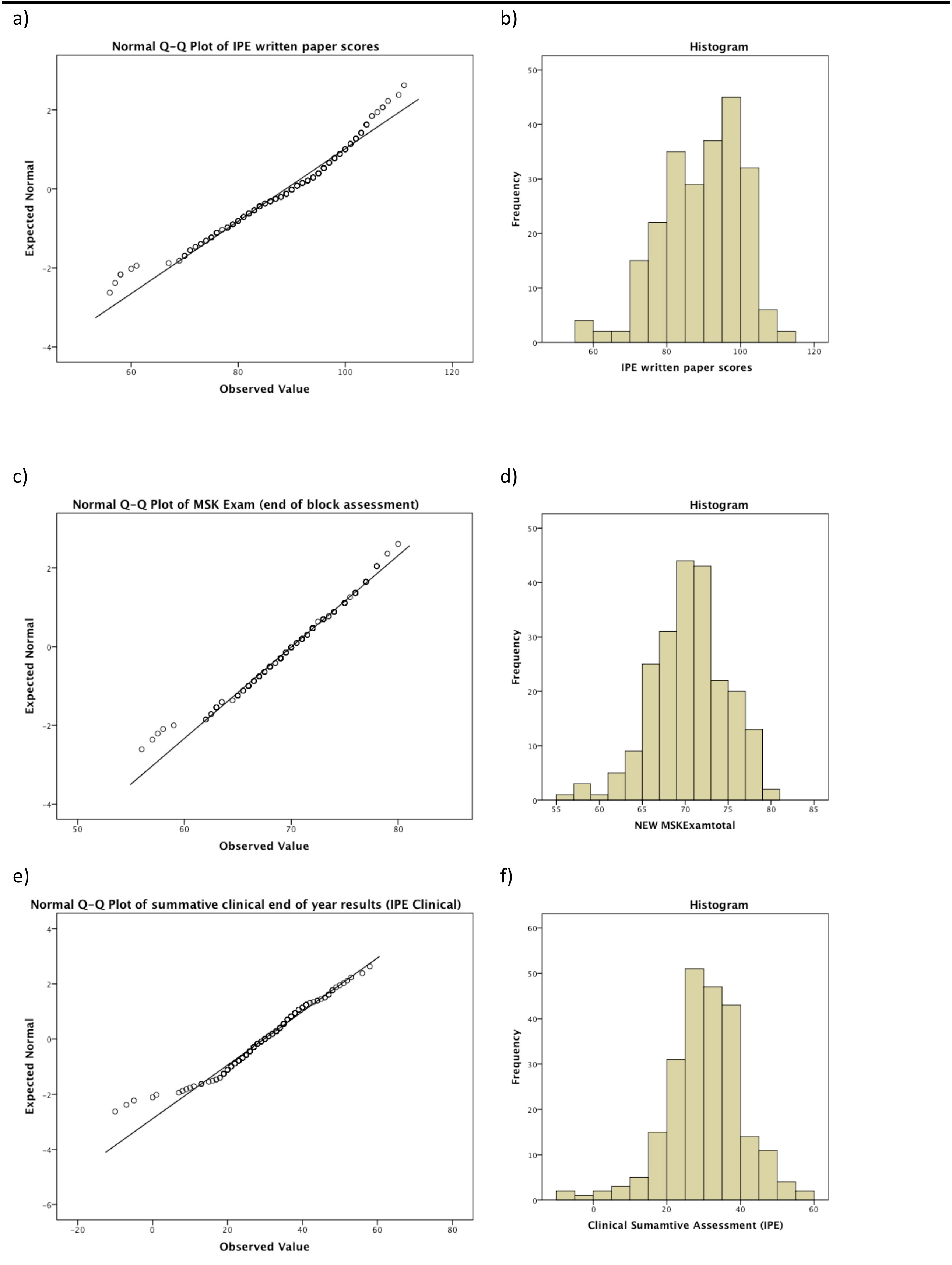


Figure 32 Descriptive statistics of assessment data from WMS. This shows QQ (a,c,e) plots and histograms for (b,d,f): summative end of year written examination performance (a,b); the MSK end of block exam (c,d); and the summative clinical exam (e,f).

6.2.5.1. *Other WMS metrics*

The diagnostic thinking inventory (DTI) scores were also normally distributed (see a,b,c).

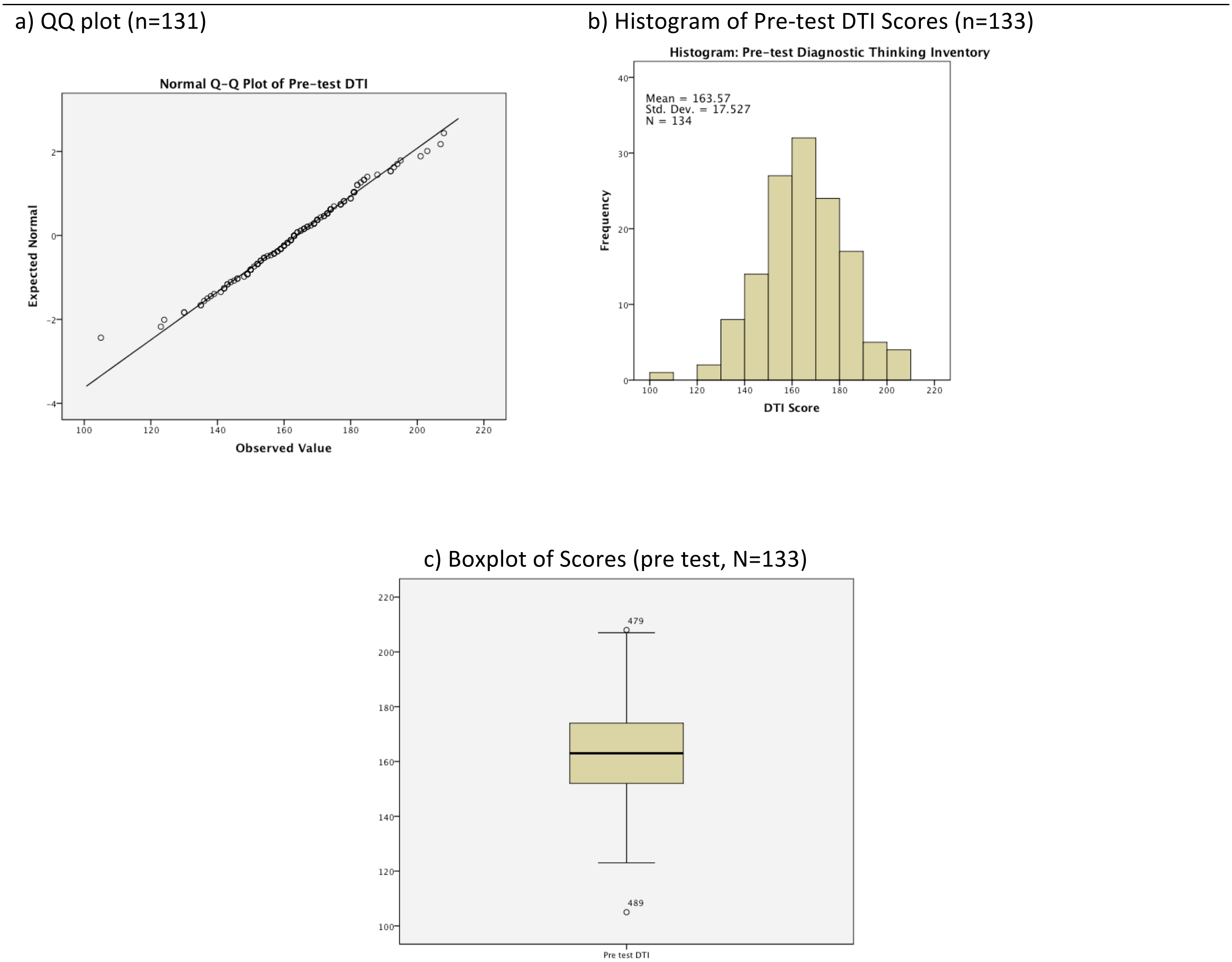


Figure 33 QQ plot (a), histogram (b) and boxplot (c) of DTI scores from 133 WMS students before completing a VP.

6.2.5.2. *Descriptive statistics by case and institution for students completing all four cases*

Boxplots for the performance in the four cases are shown below in Figure 34.

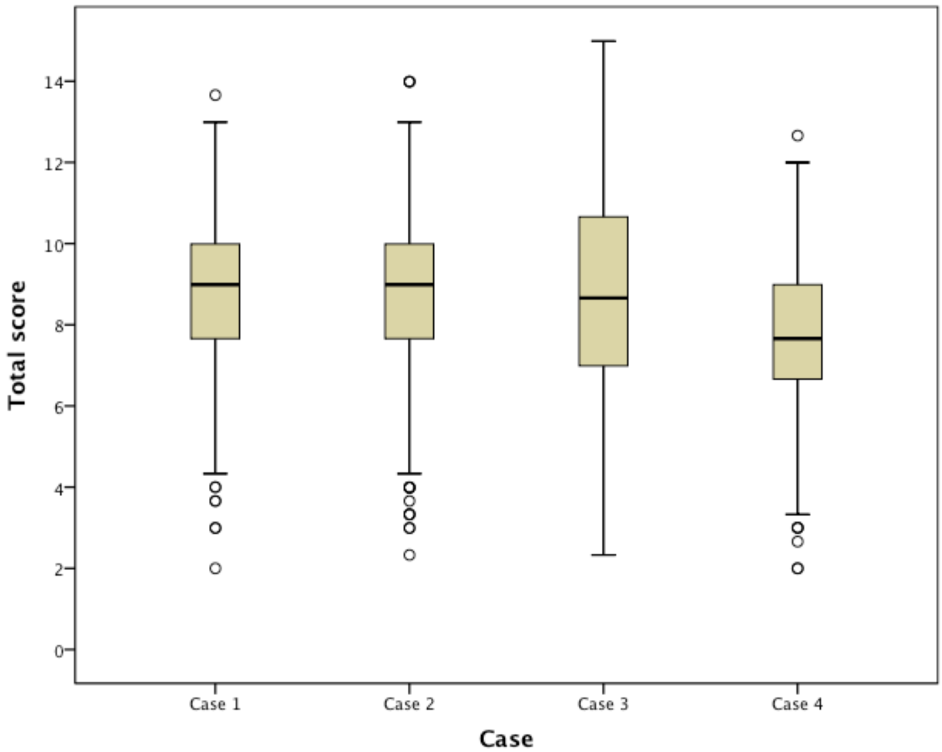


Figure 34 Boxplot for scores in the four cases (N=1773)

6.2.6. Primary Analysis: Unadjusted for two independent variables

The primary analysis was for students who completed all four cases (n=296). In the study protocol (Bateman et al., 2012a) I set an educationally significant intervention as a 5% improvement for one of the independent design variables, and a 10% improvement in VP case evaluation scores. The data analysis is an (unadjusted) independent T-test for the two groups, with the exception of the categorical variable (Bayesian reasoning, correct or incorrect). Significance is set at $p<0.05$.

For 'structured clinical reasoning instruction' (SR) there were significant differences for the number of correct diagnoses, clinical decisions, Bayesian reasoning (Table 24) and preference as a learning case (Table 25). When SR was absent, students had significantly more correct diagnoses (7.5% improvement), better clinical decisions (3.3% improvement). When SR was present, students had higher Bayes reasoning scores (7% improvement). Overall there was an improvement seen in the overall score 0.29, or 1.9% (95% CI 0.05 to 0.53) when structured clinical reasoning was not present. The only item that met the pre-determined improvement in performance of exceeding a 5% difference was the Bayesian reasoning decision (7% difference).

For 'branching' as a design variable, there were no significant differences seen for performance any marker of VP performance, or student evaluation (Table 24). Students did not prefer branching either for the case being more realistic, or a better learning experience (Table 25).

Of the 296 students who completed all four cases, 193 completed an evaluation expressing case preference (65.2%). Students were asked to indicate which case was the best for learning and which case was the most realistic. Students were asked to specify a case (1-4), and the analysis is based on the case design they had for that case. The significant finding here ($p<0.001$) was the students' preference for structured clinical reasoning (70.4%) as a better learning experience, with also a non-significant preference for this as a realistic design feature (55.4% vs. 45.6%; see Table 25).

	Branching		Structured Clinical Reasoning Instruction	
	Absent (n=592)	Present (n=592)	Absent (n=592)	Present (n=592)
Score: Diagnosis (/2)	1.34	1.36	1.43	1.28
Difference	-0.2(-0.11 to 0.06)		0.15 (0.07-0.24)	
P Value	0.627		**0.000***	
Score: KFP (/8)	4.26	4.39	4.36	4.28
Difference	-0.13 (-0.29 to 0.04)		0.08 (-0.9 to 0.24)	
P Value	0.130		.348	
Score: Clinical Decisions (/4)	2.73	2.67	**2.76***	2.64
Difference	0.07 (-0.05 to 0.18)		0.13 (0.01 to 0.24)	
P Value	0.246		**0.027***	
Score: Bayes Reasoning (/1)	0.15	0.16	0.12	**0.19***
Difference	-0.02 (0.06 to 0.02)		-0.07 (-0.11 to -0.03)	
P Value[+]			**0.001***	
Total Score: (/15)	8.49	8.59	**8.68***	8.39
Difference	-0.1 (-0.34 to 0.14)		0.29 (0.05 to 0.53)	
P Value	0.404		**0.019***[*]	
EViP Evaluation (/55)	**Absent (n=360)**	**Present (n=412)**	**Absent (n=412)**	**Present (n=360)**
	44.88	45.07	44.97	44.98
Difference	-0.19 (-0.9 to 0.55)		0.01 (-0.76 to 0.74)	
P value	.609		0.978	

[*]significant at the 0.05% level.
[+]all tests shown independent T-test other than Bayes reasoning (Pearson Chi squared)

Table 24 Primary outcome measures for the four possible case designs (students who completed all four cases, n=296).

	Branching Preference		Structured clinical Reasoning Instruction	
'Best for learning'	**Absent**	**Present**	**Absent**	**Present**
Case design preference	93 (48.2%)	100 (51.8%)	57 *(29.5%)	136[*] (70.5%)
	Asymp. Sig. 0.614, Pearson Chi-square 0.254[a]		***Asymp. Sig<.0005**, Pearson Chi-square 32.337[a]	
'Most realistic' Case design preference	88 (45.6%)	105 (54.4%)	88 (45.6%)	105 (54.4%)
	Asymp. Sig= 0.349 Pearson Chi-square 0.876[a]		Asymp. Sig.=0.221 Pearson Chi-square 1.497[a]	

Significant at the 0.001 level

Table 25 Primary outcome measure: preference of case for learning and realism, Chi-square test.

The complete primary analysis for the two independent variables for all metrics is shown for structured clinical reasoning instruction in 6.2.6.1, and for branching in 6.2.6.2.

6.2.6.1. *Unadjusted secondary analyses: structured clinical reasoning skills*

The secondary analysis looked at other case factors including time spent per case, steps, and a breakdown of the individual components of the EViP evaluation. Where structured reasoning was present, students spent more time (2.4 minutes, $P<0.0005$), took more steps (9.8, $P<0.0005$), made significantly fewer correct clinical decisions (0.13, $P<0.05$). In the secondary analysis of the students EVIP evaluations, there was a significantly higher score for the ‘coaching’ domain in the evaluation in cases where structured reasoning was absent (difference 0.28, or 1.9% difference, $P<0.01$). This change was reflected in the primary analysis but did not reach conventional statistical significance ($P=0.058$), nor did it meet the predetermined 10% improvement in evaluation scores I had set as educationally significant (Bateman et al., 2012a).

	SR	N	Mean	SD	Std. Error Mean	Sig (2-tailed)	Mean Difference	Std. Error Difference	95% C.I. difference	
Primary analysis (unadjusted): Structured Clinical Reasoning Instruction (SR) (296 students, 1184 cases)										
Minutes**	Absent	592	27.54	12.42	.51	.003**	-2.37	.79	-3.92	-.81
	Present	592	29.91	14.75	.61					
Steps**	Absent	592	51.74	5.24	.22	.000**	-9.89	.36	-10.59	-9.19
	Present	592	61.64	6.93	.28					
Diagnosis**	Absent	592	1.43	.71	.03	.000**	.15	.04	.07	.24
	Present	592	1.28	.76	.03					
KFP	Absent	592	4.37	1.48	.06	.348	.08	.08	-.09	.24
	Present	592	4.29	1.39	.06					
Clinical Decisions	Absent	592	2.76	.94	.04	.027**	.13	.06	.01	.24
	Present	592	2.64	1.02	.04					
Bayes	Absent	592	.12	.33	.01	.001** +	-.07	.02	-.11	-.03
Reasoning**	Present	592	.19	.39	.02					
Total score*	Absent	592	8.68	2.15	.09	.019*	.29	.12	.05	.53
	Present	592	8.39	2.08	.09					
Evaluation		412	8.13	1.22	.06	.640	.04	.09	-.13	.22
(authenticity)		360	8.09	1.27	.07					
Evaluation	Absent	412	15.94	2.07	.10	.117	-.24	.16	-.55	.06
(professionalism)	Present	360	16.19	2.28	.12					
Evaluation	Absent	412	12.66	1.74	.09	.055	.23	.12	-.01	.47
(coaching)	Present	360	12.43	1.60	.08					
Evaluation	Absent	412	8.24	1.12	.05	.972	.00	.08	-.16	.15
(Learning)	Present	360	8.24	1.11	.06					
Evaluation (Total)	Absent	412	44.97	5.14	.26	.978	-.01	.38	-.76	.74
	Present	360	44.98	5.36	.28					
Secondary analysis (unadjusted): Structured Clinical Reasoning Instruction. 572 students, 1773 cases										
Minutes**	Absent	907	27.32	12.60	.42	.000**	-2.72	.65	-3.99	-1.45
	Present	865	30.04	14.62	.50					
Log(10)Minutes**	Absent	907	1.396	.188	.006	.000	-.038	.009	-.0562	-.0207
	Present	865	1.4345	.193	.006					
Steps**	Absent	907	51.32	5.23	.17	.000**	-9.79	.29	-10.35	-9.23
	Present	866	61.11	6.72	.23					
KFP	Absent	907	4.30	1.48	.05	.373	.03	.03	-.04	.10
	Present	866	4.29	1.43	.05					
Diagnosis	Absent	907	1.43	.705	.023	.849	.01	.07	-.12	.15
	Present	866	1.40	.729	.025					
Clinical Decisions*	Absent	907	2.70	.97	.032	.027*	.10	.05	.01	.20
	Present	866	2.59	1.02	.035					
Bayes	Absent	907	.12	.32	.011	.001*+	-.06	.02	-.10	-.03
Reasoning**	Present	866	.18	.39	.013					
Total score	Absent	907	8.55	2.15	.07	.405	.08	.10	-.11	.28
	Present	866	8.47	2.10	.07					
Evaluation	Absent	598	8.14	1.27	.05	.154	.11	.07	-.04	.25
(authenticity)	Present	540	8.04	1.25	.05					
Evaluation	Absent	598	16.04	2.13	.09	.724	-.05	.13	-.30	.21
(professionalism)	Present	540	16.08	2.27	.10					
Evaluation	Absent	598	12.68	1.81	.07	.009*	.27	.10	.07	.47
(coaching)**	Present	540	12.41	1.70	.07					
Evaluation	Absent	598	8.26	1.18	.05	.267	.08	.07	-.06	.22
(Learning)	Present	540	8.17	1.23	.05					
Evaluation (Total)	Absent	598	45.14	5.36	.22	.180	.43	.32	-.20	1.06
	Present	540	44.70	5.50	.24					

*significant at p<0.05. **significant at p<0.01
+Pearson Chi square, see table 27.

Table 26 Primary (students who completed all four cases) and secondary analysis (all completed cases) comparing structured clinical reasoning instruction (SR): present or absent.

6.2.6.2. *Unadjusted secondary analysis: branching*

For branching case designs, the results are shown in Table 27. The only significant difference was the of the secondary analysis to the was the increased number of steps taken (more steps where branching was absent, average 1.02 per case, P<0.01). These extra steps represent students being redirected back onto the main spine of the case in the linear scenarios.

Primary analysis (unadjusted): Branching (296 students, 1184 cases)										
BRANCHING		N	Mean	Std. Devia'n	Std. Error Mean	Sig (2-tailed)	Mean Difference	Std. Error Difference	95% C.I. difference	
Minutes	Absent	592	28.34	13.07	.54	.330	-.77	.80	-2.33	.79
	Present	592	29.11	14.26	.59					
Steps**	Absent	592	57.29	7.98	.33	.010*	1.18	.46	.28	2.08
	Present	592	56.10	7.75	.32					
Diagnosis	Absent	592	1.34	.76	.03	.627	-.02	.04	-.11	.06
	Present	592	1.36	.72	.03					
KFP	Absent	592	4.26	1.38	.06	.130	-.13	.08	-.29	.04
	Present	592	4.39	1.48	.06					
Clinical Decisions	Absent	592	2.73	1.00	.04	.246	.07	.06	-.05	.18
	Present	592	2.67	.97	.04					
Bayes Reasoning	Absent	592	.15	.35	.01	.370[+]	-.02	.02	-.06	.02
	Present	592	.16	.37	.02					
Total score	Absent	592	8.49	2.11	.09	.404	-.10	.12	-.34	.14
	Present	592	8.59	2.12	.09					
Evaluation (authenticity)	Absent	360	8.06	1.24	.07	.309	-.09	.09	-.27	.08
	Present	412	8.15	1.25	.06					
Evaluation (professionalism)	Absent	360	16.01	2.21	.12	.567	-.09	.16	-.40	.22
	Present	412	16.10	2.14	.10					
Evaluation (coaching)	Absent	360	12.58	1.64	.09	.694	.05	.12	-.19	.28
	Present	412	12.53	1.72	.08					
Evaluation (Learning)	Absent	360	8.22	1.09	.06	.723	-.03	.08	-.19	.13
	Present	412	8.25	1.13	.06					
Evaluation (Total)	Absent	360	44.88	5.15	.27	.609	-.19	.38	-.94	.55
	Present	412	45.07	5.33	.27					
Secondary analysis (unadjusted): Branching. 572 students, 1773 cases										
Minutes	Absent	874	28.44	13.21	.45	.534	-.40	.65	-1.68	.87
	Present	899	28.84	14.13	.47					
Steps	Absent	874	56.62	7.74	.26	.006**	1.02	.37	.30	1.74
	Present	899	55.60	7.72	.26					
KFP	Absent	874	1.40	.74	.03	.397	-.03	.03	-.10	.04
	Present	899	1.43	.69	.02					
Diagnosis	Absent	874	4.23	1.43	.05	.095	-.12	.07	-.25	.02
	Present	899	4.35	1.48	.05					
Clinical Decisions	Absent	874	2.67	1.00	.03	.281	.05	.05	-.04	.14
	Present	899	2.62	.99	.03					
Bayes Reasoning	Absent	874	.14	.35	.01	.312[+]	-.02	.02	-.05	.02
	Present	899	.16	.37	.01					
Total score	Absent	874	8.45	2.15	.07	.260	-.11	.10	-.31	.08
	Present	899	8.57	2.10	.07					
Evaluation (authenticity)	Absent	536	8.06	1.24	.05	.394	-.063	.074	-.209	.082
	Present	618	8.12	1.27	.05					
Evaluation (professionalism)	Absent	536	16.02	2.23	.10	.609	-.07	.14	-.324	.19
	Present	618	16.09	2.16	.09					
Evaluation (coaching)	Absent	536	12.59	1.72	.07	.467	.08	.104	-.13	.28
	Present	618	12.51	1.80	.07					
Evaluation (Learning)	Absent	536	8.23	1.20	.05	.825	.02	.071	-.12	.16
	Present	618	8.21	1.21	.05					
Evaluation (Total)	Absent	536	44.90	5.32	.23	.837	-.07	.32	-.70	.57
	Present	618	44.96	5.52	.22					

[+]Pearson Chi square, see table 27 *significant at p<0.05 **significant at p<0.01

Table 27 Primary (students who completed all four cases) and secondary analysis (all completed cases) comparing branching: present or absent.

For clarity I present the comparison between the Bayesian reasoning for the two design variables. This supports that the difference in Bayesian reasoning in the 'SR' variable was significant (P<0.001) however for the branching it was not (P=0.377).

A.) Branching

		Bayes Reasoning		
		Incorrect	Correct	Total
branching	Absent	506	86	592
	Present	495	97	592
Total		1001	183	1184

Mean Bayes Score: Absent= 15%, Present =16%

Pearson Chi-square: .78, d.f.1, Asyp Sig (2-sided)=0.377

B.) Structured clinical reasoning instruction

		Bayes Reasoning		
		Incorrect	Correct	Total
Structured reasoning	Absent	521	71	592
	Present	480	112	592
Total		1001	183	1184

Mean: Absent= 12% Present =19%

Pearson Chi-square: 8.9, d.f.1, Asyp Sig (2-sided)=<0.001

Table 28 Bayes reasoning performance for the two independent design variables- Pearson's Chi Squared

6.2.7. Planned Primary analysis: ANCOVA for the VP scores

Analysis of covariance (ANCOVA) was part of pre-determined data analysis plan (see 6.1.12, p.154), to determine the main effects for the two independent variables (branching and structured clinical reasoning instruction), and any interaction effects between the design variables. This is part of the primary analysis from the 296 students completing 1184 VPs. Each student completed one of the case designs (1a, 1b, 2a, 2b). ANCOVA allows the adjustment for other fixed factors, or categorical predictor variables. The three variables I used were gender (M=117, F=179), recruitment centre (WMS n=136, UBMS n=116, KMS n=44) and the VP case (case 1-4). The ANVOCA was performed for the overall performance in the cases, and for the total evaluation scores.

6.2.7.1. *Suitability for ANCOVA analysis*

The error variance was equal across the groups supporting the use of ANCOVA as defined by a Levine's test showing no significant differences in the equality of the error variance (Levine's test: F=1.215, df189, df 2 1094, Sig=0.092). The results are shown in Table 29 (p.192).

Results for 1184 completed VPs by 296 students

A.) ANCOVA for main effects and between-subjects effects for performance in the case (score /15)

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	1128.632[a]	89	12.681	3.332	.000	.213
Intercept	33896.166	1	33896.166	8907.438	.000	.891
Main effects						
Structured Reasoning (SR)	2.277	1	2.277	.598	.439	.001
Branching	2.784	1	2.784	.732	.393	.001
Gender	1.531	1	1.531	.402	.526	.000
Recruiting centre	162.925	2	81.462	21.407	.000	.038
Case number (1-4)	67.809	3	22.603	5.940	.001	.016
Interaction effects						
SR * Branching	1.112	1	1.112	.292	.589	.000
SR * Gender	.252	1	.252	.066	.797	.000
SR * recruiting centre	9.611	2	4.806	1.263	.283	.002
SR * Case	22.753	3	7.584	1.993	.113	.005
Branching * gender	1.249	1	1.249	.328	.567	.000
Branching * recruiting centre	1.277	2	.638	.168	.846	.000
Branching * Case (1-4)	17.946	3	5.982	1.572	.194	.004
Error	4163.083	1094	3.805			
Total	91592.522	1184				
Corrected Total	5291.715	1183				

B.) Adjusted population means for independent variables

Independent Variable		Score per VP (/15)	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Structured clinical reasoning instruction					
	Absent	8.641	.109	8.428	8.855
	Present	8.650 [a]	.113	8.428	8.872
Branching					
	Absent	8.553 [a]	.113	8.332	8.774
	Present	8.738 [a]	.110	8.523	8.953
Structured reasoning * Branching					
SR	*Branching*				
Absent	*Absent*	8.602 [a]	.166	8.276	8.927
Absent	*Present*	8.683 [a]	.140	8.409	8.957
Present	*Absent*	8.503 [a]	.151	8.206	8.800
Present	*Present*	8.791 [a]	.168	8.462	9.120
Fixed Factors (main effect displayed)					
Gender	Male	8.696 [a]	.129	8.443	8.948
	Female	8.594 [a]	.088	8.421	8.766
Recruiting Centre	WMS	8.250 [a]	.108	8.037	8.462
	UBMS	9.581 [a] *	.137	9.313	9.849
	KMS	8.157 [a]	.154	7.854	8.460
VP case number	1	8.995[a]	0.143	8.715	9.275
	2	8.973[a]	0.143	8.693	9.253
	3	8.751[a]	0.143	8.471	9.031
	4	7.605[a] **	0.214	7.186	8.024

a. Based on modified population means

*P<0.001 on post hoc (Tukey) tests as significantly higher than the other two centres

**p<0.001 on post hoc (Tukey) test as significantly lower than other case designs

Table 29 ANCOVA for the five fixed variables showing: A.) Main and interaction effects of the variables on case scores (N=1184 cases); B.) Adjusted population means (Gender, recruitment centre, structured reasoning, branching, VP case number).

The main effect for the two independent design variables on case performance was not significant (structured reasoning $F=.47$ $p=0.491$; branching $F=0.73$, $p=0.393$). Furthermore there was no significant interaction effect between the two design variables ($F=0.29$, $p=0.59$). This suggests that branching does not influence the effectiveness of clinical reasoning instruction, and *vice versa*. There were no significant interaction effects between the two design variables and gender, recruitment centre, or case being completed (all $p>0.1$)

There was a significant independent main effect for two of the fixed variables, the recruitment centre ($F=21.40$, $p<0.0005$, partial Eta squared 0.038, small effect size) and case ($F=5.94$, $p<0.001$, partial Eta squared=0.016, small effect size). The effect sizes for partial Eta squared used are as follows: small >0.01; Medium >0.06; Large >0.14 (Tabachnick and Fidell, 2007).

6.2.8. Planned Primary analysis: ANCOVA for the evaluation data.

The statistics for the completed evaluations are shown below for the 296 students who completed four VPs. Evaluations were not compulsory, and the completion rates used in the ANCOVA are shown below in Table 30.

		Value Label	N (
Design structured reasoning	1	Absent	412
	2	Present	360
Design: branching	1	Absent	360
	2	Present	412
University	1	WMS	365
	2	UBMS	289
	3	KMS	118
Gender	1	Male	292
	2	Female	480
Case	1	Case 1	226
	2	Case 2	206
	3	Case 3	193
	4	Case 4	147

Table 30 Between-Subjects Factors for the 772 completed VP evaluation scores (students completing all four cases, N=296, N=1184 cases)

The error variance for the evaluation scores was equal across the groups (Levine's test non significant: F=1.21, df 1 87, df 2 684, p=0.103).

There was no significant difference between the adjusted population means for the two groups for either branching (P=0.89) or structured clinical reasoning instruction (p=0.26), see Table 31. There was no interaction effect seen between the two independent variables in terms of student evaluations (P=0.69).

Complete evaluations from 1184 completed VPs, (n=296 students).

A.) ANCOVA for main effects and between-subjects effects for evaluation (/55)

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	3904.245[a]	87	44.876	1.777	.000	.186
Intercept	740385.031	1	740385.031	29319.814	.000	.977
SR	32.751	1	32.751	1.297	.255	.002
Branching	.509	1	.509	.020	.887	.000
UNIVERSITY	838.870	2	419.435	16.610	.000	.047
GENDER	84.291	1	84.291	3.338	.068	.005
Case	6.889	3	2.296	.091	.965	.000
SR * Branching	4.119	1	4.119	.163	.686	.000
Total	1566545.000	772				
Corrected Total	20974.622	778				

a. R Squared = .186 (Adjusted R Squared = .081)

B.) Adjusted population means for independent variables

Independent Variable		**Evaluation Score**	**Std. Error**	**95% Confidence Interval Lower Bound**	**Upper Bound**
Structured clinical reasoning instruction					
	Absent	45.27[a]	.340	44.600	45.94
	Present	45.46[a]	.363	44.746	46.17
Branching					
	Absent	45.34[a]	.378	44.60	46.08
	Present	45.39[a]	.324	44.75	46.02
Structured reasoning * Branching					
Structured reasoning	*Branching*				
Absent	*Absent*	45.32[a]	0.52	44.29	46.35
Absent	*Present*	45.17[a]	0.44	44.31	46.02
Present	*Absent*	45.31[a]	0.55	44.24	46.38
Present	*Present*	45.56[a]	0.48	44.61	46.50
Fixed Factors (main effect displayed)					
Gender	Male	45.73[a]	.40	44.942	46.518
	Female	45.00[a]	.29	44.420	45.575
Recruiting Centre	WMS	43.69[a]	.32	43.050	44.325
	UBMS	46.99[a]	.38	46.243	47.742
	KMS	45.51[a]	.53	44.473	46.544
VP case number	1	45.30[a]	.438	44.444	46.163
	2	45.40[a]	.485	44.453	46.356
	3	45.17[a]	.440	44.303	46.030
	4	45.53[a]	.679	44.196	46.862

Table 31 ANCOVA for the five fixed variables showing: A.) Main and interaction effects of the variables on evaluation scores (N=1184 cases); B.) Adjusted population means (Gender, recruitment centre, structured reasoning, branching, VP case number).

6.2.9. Primary analysis: Unadjusted ANOVA of the four case designs against VP performance metrics

To further explore the data I have performed an analysis of variance for the four case designs against each of the secondary outcome measures. I then used post-hoc multiple comparison tests (Tukey) to look for any suggestions of independent or interaction effects between the four case designs that were not revealed by the previous analysis. The results are presented for students who have completed all four cases in Table 32. There were significant differences for time spent on the case (P=0.019, small effect size), and diagnoses made (p<0.01, small effect size), using the previously defined effect sizes (>0.01; Medium >0.06; Large >0.14).

Case Design		N	Mean	Std. Deviation	Std. Error	95% C.I. Lower	95% C.I. Upper	F	Significance	Partial Eta Squared
Time Spent (Minutes)	1A	296	27.02	11.51	.67	25.70	28.34	3.324	**.019**	**.01**
	1B	296	28.06	13.27	.77	26.55	29.58	.007)		(small effect size)
	2A	296	29.66	14.36	.84	28.01	31.30			
	2B	296	30.16	15.14	.88	28.43	31.89			
	Total	1184	28.73	13.68	.40	27.95	29.51			
Diagnosis	1A	296	1.43	.72	.04	1.35	1.51	4.393	**.004**	**.01**
	1B	296	1.43	.71	.04	1.35	1.51			(small effect size)
	2A	296	1.26	.78	.05	1.17	1.35			
	2B	296	1.30	.74	.04	1.21	1.38			
	Total	1184	1.35	.74	.02	1.31	1.40			
KFP	1A	296	4.25	1.39	.08	4.10	4.41	1.534	.204	N/A
	1B	296	4.48	1.56	.09	4.30	4.66			
	2A	296	4.28	1.38	.08	4.12	4.43			
	2B	296	4.30	1.39	.08	4.14	4.46			
	Total	1184	4.33	1.43	.04	4.25	4.41			
Clinical Decisions	1A	296	2.83	.94	.05	2.72	2.94	2.578	.052	N/A
	1B	296	2.69	.93	.05	2.59	2.80			
	2A	296	2.64	1.05	.06	2.52	2.76			
	2B	296	2.64	1.00	.06	2.52	2.75			
	Total	1184	2.70	.98	.03	2.64	2.76			
Total score	1A	296	8.63	2.11	.12	8.39	8.87	2.068	.103	N/A
	1B	296	8.73	2.18	.13	8.48	8.98			
	2A	296	8.34	2.11	.12	8.10	8.58			
	2B	296	8.45	2.05	.12	8.21	8.68			
	Total	1184	8.54	2.12	.06	8.42	8.66			

Levene Statistic:
Minutes, 2.150, DF1 3, DF2 1180, Sig .092; Diagnosis 1.446, DF1 3, DF2 1179 Sig .228; KFP 2.288, DF1 3, DF2 1179,Sig .077; Clinical Decisions 3.572, DF1 3, DF2 1179 Sig .014; Bayes Reasoning 16.210 DF1 3, DF2 1179, Sig .000; Total score .419, DF1 3, DF 2 1179, Sig .740

Table 32 Results from ANOVA looking a VP case design against performance metrics.

When evaluating these with multiple comparisons the differences for the two properties that were highlighted as significant (time spent and diagnoses) are shown below in Table 33.

Dependent Variable	(I) Case Design	(J) Case Design	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	Upper Bound
Time spent (Minutes)	1A	1B	-1.04	1.12	0.789	-3.93	1.84
		2A	-2.64	1.12	0.088	-5.52	0.25
		2B	-3.14*	1.12	0.027	-6.03	-0.26
Diagnosis	1A	1B	0.00	0.06	1.000	-0.16	0.15
		2A	0.17*	0.06	0.023	0.02	0.33
		2B	0.13	0.06	0.130	-0.02	0.29
	1B	1A	0.00	0.06	1.000	-0.15	0.16
		2A	0.17*	0.06	0.022	0.02	0.33
		2B	0.13	0.06	0.124	-0.02	0.29

The mean difference is significant at the 0.05 level

+ There were no significant differences between design typologies for the domains not presented here (Diagnosis, KFP, total score)

Table 33 Post Hoc multiple comparisons (Tukey) for independent design variables: showing significant differences found for diagnosis, time, steps and Bayes reasoning

6.2.9.1. *Primary analysis: Unadjusted ANOVA of the four case designs against VP evaluation data.*

When evaluating the case evaluations, there were no additional findings from the ANOVA that supported any significant improvement in any of the evaluation metrics across the design typologies (all p>.13). This is shown in Table 34.

Case Design		N	Mean	Std. Deviation	Std. Error	95% C.I. Lower	95% C.I. Upper	ANOVA F	ANOVA Significance
Evaluation (authenticity)	1A	196	8.15	1.15	.08	7.99	8.31	1.124	.339
	1B	218	8.12	1.29	.09	7.95	8.29		
	2A	165	7.96	1.34	.10	7.76	8.17		
	2B	195	8.19	1.21	.09	8.03	8.36		
	Total	772	8.11	1.25	.04	8.03	8.20		
Evaluation (professionalism)	1A	196	15.95	2.00	.14	15.67	16.23	1.060	.365
	1B	218	15.94	2.14	.14	15.65	16.22		
	2A	165	16.08	2.45	.19	15.70	16.46		
	2B	195	16.27	2.124	.150	15.98	16.57		
	Total	772	16.06	2.172	.078	15.90	16.21		
Evaluation (coaching)	1A	196	12.76	1.666	.119	12.53	13.00	1.887	.130
	1B	218	12.57	1.809	.122	12.33	12.81		
	2A	165	12.35	1.589	.124	12.11	12.60		
	2B	195	12.49	1.616	.114	12.26	12.71		
	Total	772	12.55	1.682	.060	12.43	12.67		
Evaluation (Learning)	1A	196	8.26	1.050	.075	8.12	8.41	.447	.720
	1B	218	8.21	1.174	.079	8.05	8.37		
	2A	165	8.17	1.144	.089	7.99	8.35		
	2B	195	8.29	1.086	.077	8.14	8.44		
	Total	772	8.24	1.114	.040	8.16	8.32		
Evaluation (Total)	1A	196	45.13	4.785	.342	44.46	45.81	.752	.521
	1B	218	44.82	5.463	.377	44.08	45.57		
	2A	165	44.57	5.550	.433	43.71	45.42		
	2B	195	45.34	5.185	.372	44.60	46.07		
	Total	772	44.98	5.243	.190	44.61	45.35		

Levene Statistic satisfactory for all domains: Evaluation (authenticity) .145, DF1 3, DF2 777,Sig .933; Evaluation (professionalism) 1.765, DF1 3 DF2 777, Sig .152; Evaluation (coaching) .349, DF1 3, DF2 777, Sig .790; Evaluation (Learning) .594, DF1 3, DF2 777, Sig .619, Evaluation (Total) .648, DF1 3, DF2 760, Sig .584

Table 34 ANOVA by case design for students who completed all four VP cases- no significant differences seen

6.2.10. Secondary analysis

I have already presented the lack of significant differences in the secondary analysis for all cases completed (n=1773 VPs) compared to students completing all four cases (n=1184 VPs). This is unsurprising given the lack of a significant association I have shown between: (1) scores between those dropping out of the study; (2) evaluation completion rates by score in VP; and (3) future summative assessment achievement in those that consent or do not consent to participate. In this secondary analysis I present data relating to the impact of the institution and extent of curricular integration.

6.2.10.1. Impact of the institution

This analysis was conducted due to some of the findings from the ANCOVA suggesting institution is an important fixed factor in student performance. The boxplots of performance by institution across the cases are shown in Figure 35.

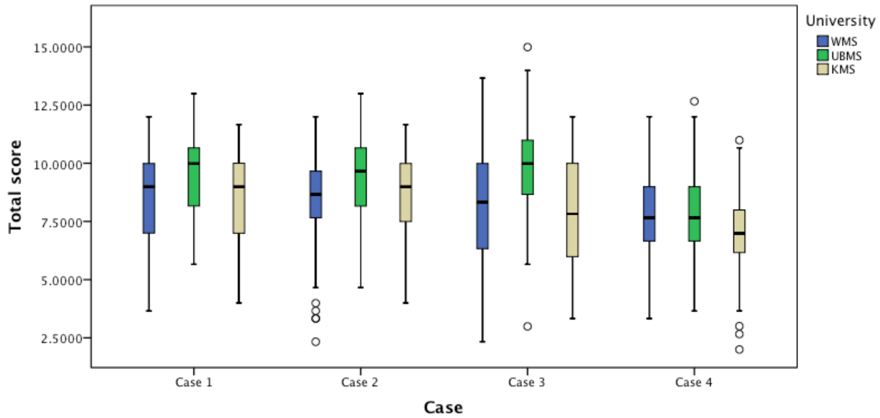


Figure 35 Box plot for VP performance in the four cases (unadjusted) for the three schools across the four cases (n=1773)

The ANOVA with post-hoc multiple comparisons by institution show the UBMS students have higher results for score, evaluation and time spent per case than WMS (P<0.001) and KMS (P<0.001). There were no significant differences between WMS and KMS for any outcome (see Table 35).

		N	Mean	Std. Deviation	95% Confidence Interval for Mean		F	Sig.
					Lower Bound	Upper Bound		
Total score	WMS	544	8.19	2.01	8.02	8.35	33.83	P<0.001
	UBMS	464	9.16	2.05	8.97	9.36		
	KMS	176	8.07	2.24	7.73	8.40		
	Total	1184	8.54	2.11	8.42	8.66		
Evaluation (Total)	WMS	367	43.79	5.45	43.23	44.35	21.07	P<0.001
	UBMS	289	46.39	4.97	45.82	46.97		
	KMS	118	45.07	4.22	44.30	45.84		
	Total	774	44.96	5.23	44.59	45.33		
Minutes	WMS	544	27.04	11.57	26.08	28.00	31.56	P<0.001
	UBMS	464	32.57	16.45	31.03	34.10		
	KMS	176	24.52	9.27	23.14	25.90		
	Total	1184	28.75	13.70	27.97	29.53		

Post Hoc tests (Tukey): multiple comparisons show UBMS significantly (p<0.01) higher for both WMS and KMS for or all three metrics; no significant differences between WMS and KMS (P all>0.4)

Table 35 ANOVA of VP performance, evaluation scores, and time spent by institution.

6.2.11. VP scores and other metrics

I have analysed two correlation matrices for VP scores. The first compares performance from students who completed all four cases (296). For the 140 WMS students, I have compared their performance in the summative written and clinical examinations. The secondary analysis compares the mean scores for the 572 students who completed at least one VP case. I have used Pearson's Correlation 'R'. The effect size for correlation coefficient is determined by Cohen's criteria for Pearson Correlation (1992): small effect size=0.1-0.23; medium 0.24-0.36; large >0.37 or greater.

6.2.11.1. Correlations for the students completing all four VPs

The summative assessment scores available are an end of year written paper (Intermediate Professional Exam, or IPE), and end of year clinical examination (IPE Clinical), and an end of block musculoskeletal written (MSK written) and MSK three station OSCE (Objective Structured Clinical Examination). The correlations described below are shown in Table 36.For performance in the end of year summative written exam, the three strongest correlations were: (1) KFP scores across the four cases (Pearson's R=0.42, P<.0005, large effect size); (2) Total VP marks (R=0.39, P<0.0005, large effect size); (3) the MSK two station OSCE (R=0.20, P<0.05, small-medium effect size). There are also two metrics that correlate significantly with the MSK OSCE score: the total VP score (R=0.20, small-medium effect size, P<0.05), and IPE written paper (R=0.19, small-medium effect size, P<.05). No measured metrics correlated with the end of block summative clinical examination, although the KFP score in the cases approached statistical significance (p=0.053). There was a small number of missing summative assessment scores, 3/140 for the IPE examinations, and 11/140 for the MSK examinations, see Table 36.

Correlations

		Total	KFP	Time	Steps	Evaluation	MSK written†	MSK OSCE†	IPE Written†	IPE Clinical†
VP Total	Pearson Correlation	1	.845**	.162**	.071	.192**	.014	.233**	.377**	.157
	Sig. (2-tailed)		.000	.005	.226	.002	.872	.009	.000	.070
	N	296	296	296	296	268	126	126	134	134
KFP	Pearson Correlation	.845**	1	.161**	.158**	.128*	.114	.098	.424**	.169
	Sig. (2-tailed)	.000		.006	.006	.036	.206	.276	.000	.051
	N	296	296	296	296	268	126	126	134	134
TIme	Pearson Correlation	.162**	.161**	1	.081	.167**	.003	-.038	-.202*	.139
	Sig. (2-tailed)	.005	.006		.162	.006	.974	.677	.020	.110
	N	296	296	296	296	268	126	126	134	134
Steps	Pearson Correlation	.071	.158**	.081	1	.044	.087	.072	.013	.182*
	Sig. (2-tailed)	.226	.006	.162		.478	.330	.425	.879	.035
	N	296	296	296	296	268	126	126	134	134
Evaluations	Pearson Correlation	.192**	.128*	.167**	.044	1	.025	.131	.059	-.014
	Sig. (2-tailed)	.002	.036	.006	.478		.785	.158	.513	.873
	N	268	268	268	268	268	118	118	126	126
MSK Written†	Pearson Correlation	.014	.114	.003	.087	.025	1	-.075	.364**	.034
	Sig. (2-tailed)	.872	.206	.974	.330	.785		.401	.000	.699
	N	126	126	125	126	118	129	129	129	129
MSK OSCE†	Pearson Correlation	.233**	.098	-.038	.072	.131	-.075	1	.191*	.027
	Sig. (2-tailed)	.009	.276	.677	.425	.158	.401		.031	.759
	N	126	126	125	126	118	129	129	129	129
IPE Written†	Pearson Correlation	.377**	.424**	-.202*	.013	.059	.364**	.191*	1	.142
	Sig. (2-tailed)	.000	.000	.020	.879	.513	.000	.031		.100
	N	134	134	133	134	126	129	129	136	136
IPE Clinical†	Pearson Correlation	.157	.169	.139	.182*	-.014	.034	.027	.142	1
	Sig. (2-tailed)	.070	.051	.110	.035	.873	.699	.759	.100	
	N	134	134	133	134	126	129	129	136	136

†137/140 WMS IPE assessment results were unavailable (3 missing, 2.1%)
129 of 140 MSK examination results were available (11 missing, 7.9%)

Table 36 Correlation matrix for WMS students who completed all four VPs

By dividing the WMS cohort who completed all four cases (N=136) into four quartiles it is possible to explore their performance on the IPE written end of year paper. Results were unavailable for 11/140 students for the MSK exam and 3/140 students for the written and clinical exams.

Students in the first quartile performed significantly worse than those in all the other quartiles, with mean written paper marks increasing across the four quartiles (quartile 1, 82.5; quartile 2, 89.5; quartile 3 90.9; quartile 4, 93.1). This is shown in Table 37, with a boxplot of the four quartiles performance in Figure 36.

Quartile for Mean VP score	N	Mean IPE written	Std. Deviation	95% C.I. Lower	95% C.I. Upper	Minimum	Maximum
1 [0-7.1)	34	82.47	12.016	78.28	86.66	56	104
2 (7.10-8.37)	34	89.52	10.61	85.75	93.28	61	107
3 (8.37-9.08)	34	90.56	11.253	86.63	94.48	58	106
4 (>9.08)	34	93.15	9.504	89.83	96.52	67	110
Total	136	88.99	11.498	87.03	90.94	56	110

ANOVA: Equality of variance satisfactory (Levene's test p=0.8). Significant difference between groups (p<0.001)
Tukey post hoc test: Quartile 1 significantly lower than Quartiles 2 (p<0.05), 3 (p<0.01), and 4 (p<0.001), no other significant differences

Table 37 Four quartiles in performance from WMS and ANOVA showing significant differences between the quartiles

These results can also be presented as a boxplot below.

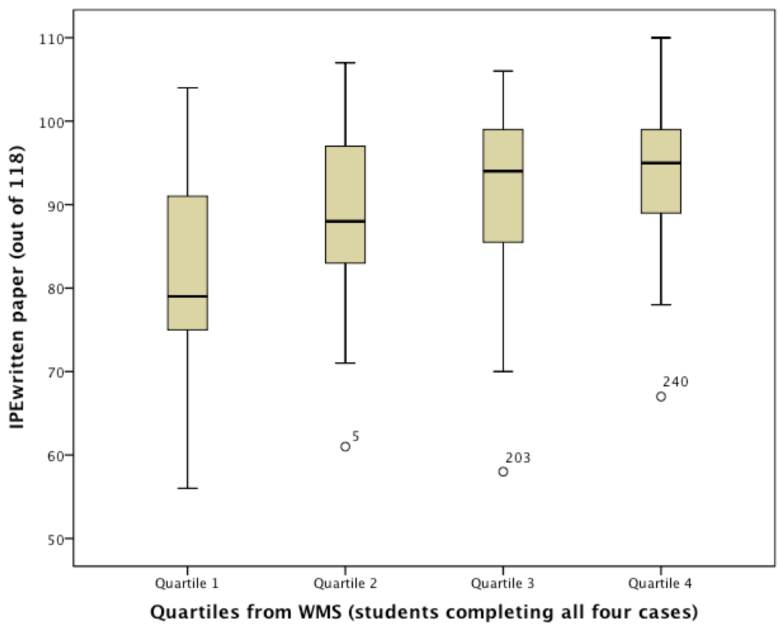


Figure 36 Boxplots relating scores across all four VP cases to the students written exam results.

6.2.11.2. All students who completed at least one case

In Table 38 I present a Pearson correlation of all the 572 students taking part in the research, presenting mean performance metrics for every student.

Mean VP scores from over 200 WMS students correlated with the MSK OSCE (R=0.15, p<0.04), the IPE written examination(R=0.34, p<0.001), and also the IPE clinical examination ($r=0.175$, $p<0.01$). Evaluation scores correlate with time spent per case ($R=0.15$, $p<0.001$, small effect size) and VP score ($R=0.15$, $p<0.001$, small effect size). The strongest positive correlation with VP score was the pre-test diagnostic thinking inventory score.

There was one item that correlated negatively with summative written exam, and this was the time spent per case ($r=-0.21$, $p<0.01$, small-medium effect size). The time spent per case by individual did correlate positively with Bayesian reasoning ($R=.102$), VP performance score ($R=.124$), and student evaluation (0.153).

		DTI	Time	Steps	KFP	Bayes	VP total	Mean evaluation	MSK OSCE	MSK Written	IPE written	IPE Clinical
DTI Pre	Pearson Correlation	1	-.205*	.056	-.019	-.087	-.146	.351**	.073	.138	.027	.100
	Sig. (2-tailed)		.020	.531	.830	.331	.102	.000	.415	.122	.754	.248
	N	134	127	127	127	127	127	134	128	128	134	134
Time	Pearson Correlation	-.205*	1	.120**	.001	.102*	.124**	.153**	-.060	-.074	-.209**	.060
	Sig. (2-tailed)	.020		.004	.972	.015	.003	.001	.386	.288	.002	.379
	N	129	572	572	572	572	572	470	210	210	215	215
Steps	Pearson Correlation	.056	.120**	1	.011	.130**	.053	.014	.016	-.050	-.012	.053
	Sig. (2-tailed)	.531	.004		.800	.002	.207	.766	.818	.470	.866	.441
	N	127	572	572	572	572	572	468	207	207	215	215
KFP	Pearson Correlation	-.087	.102*	.130**	.051	1	.850**	.091*	.102	.098	.334**	.201**
	Sig. (2-tailed)	.331	.015	.002	.219		.000	.050	.144	.160	.000	.003
	N	127	572	572	572	572	572	468	207	207	215	215
Bayes	Pearson Correlation	-.019	.001	.011	1	.051	.135**	-.029	.007	-.009	-.075	-.041
	Sig. (2-tailed)	.830	.972	.800		.219	.001	.524	.924	.896	.270	.551
	N	127	572	572	572	572	572	468	207	207	215	215
VP Total	Pearson Correlation	-.146	.124**	.053	.135**	.850**	1	.150**	.154*	.022	.344**	.175*
	Sig. (2-tailed)	.102	.003	.207	.001	.000		.001	.027	.755	.000	.010
	N	127	572	572	572	572	572	468	207	207	215	215
EViP Eval.	Pearson Correlation	.351**	.153**	.014	-.029	.091*	.150**	1	-.025	.064	-.015	.012
	Sig. (2-tailed)	.000	.001	.766	.524	.050	.001		.744	.400	.843	.866
	N	109	470	468	468	468	468	478	177	177	186	186
MSK OSCE	Pearson Correlation	.073	-.060	.016	.007	.102	.154*	-.025	1	.062	.240**	.043
	Sig. (2-tailed)	.415	.386	.818	.924	.144	.027	.744		.364	.000	.529
	N	128	210	207	207	207	207	177	218	218	218	218
MSK Written	Pearson Correlation	.138	-.074	-.050	-.009	.098	.022	.064	.062	1	.340**	.041
	Sig. (2-tailed)	.122	.288	.470	.896	.160	.755	.400	.364		.000	.544
	N	128	210	207	207	207	207	177	218	218	218	218
IPE written	Pearson Correlation	.027	-.209**	-.012	-.075	.334**	.344**	-.015	.240**	.340**	1	.169*
	Sig. (2-tailed)	.754	.002	.866	.270	.000	.000	.843	.000	.000		.010
	N	134	215	215	215	215	215	186	218	218	228	228
IPE Clinical	Pearson Correlation	.100	.060	.053	-.041	.201**	.175*	.012	.043	.041	.169*	1
	Sig. (2-tailed)	.248	.379	.441	.551	.003	.010	.866	.529	.544	.010	
	N	134	215	215	215	215	215	186	218	218	228	228

Table 38 Correlation matrix showing mean results of VP metrics from all students completing at least one case (n=572), including correlations with DTI and assessment data (WMS only)

6.2.11.3. Diagnostic Thinking Inventory scores (DTI)

Of the 229 students from WMS, 161 were offered the DTI to complete pre- and post-test as part of the study. From this 161 invited to participate, 150 consented, and returned 134 completed pre-VP DTI scores (89.3%). From this 134, 127 went on to complete at least 1 VP case. In total 85 complete post-VP DTI forms were returned (56.7%), with 80 complete paired forms returned (pre- and post-test DTI) from eligible consenting students (53.3%, 80/150). Pre- and post-test DTI scores were strongly correlated ($p<0.0001$, Pearson's $R=0.69$, large effect), however there was a non-significant small improvement of 0.8 only post-test (paired sample t-test, $p=0.712$, mean increase 0.88), see Table 39.

		N=166 students offered DTI from 229 WMS students	*Mean (SD) [95% C.I.*
WMS Students offered DTI		161	
Offered DTI and Consented to Study		150/161 (93.2%)	
	Completed pre-VP DTI (%)	134/150 (89.3%)	163.59 (17.6), 95% CI 160.5-166.5
	Completed Post-VP DTI (%)	85/150 (56.7%)	163.56 (18.0), 95% CI 160.3-166.6
Paired Sample Statistics (N=82 pairs)			
	DTI Pre	82 (54.6%)	163.59 (17.48)
	DTI Post	82 (54.6%)	164.16 (18.1)
	Paired samples correlation: Pearson's R=0.695, p<0.001		

	Paired Differences Pre-post intervention							
				95% Confidence Interval of the Difference				
	Mean	Std. Deviation	Std. Error Mean	Lower	Upper	t	df	Sig. (2-tailed)
Paired Samples Test	-.878	13.8	1.5	-3.9	2.2	-.575	81	.567

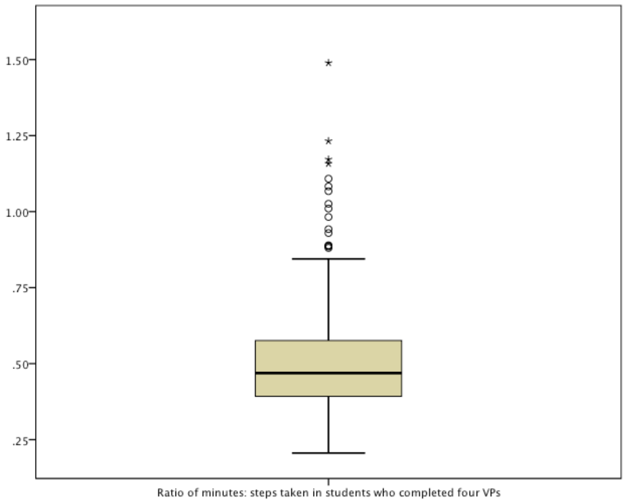
Table 39 Descriptive statistics for DTI and paired sample T-test showing a small but non-significant improvement in DTI scores post intervention.

I have already described that DTI scores did not correlate with any metrics measured within the cases, or with any university summative assessment for either mean case performance for all WMS students (see Table 38) or for the subgroup who completed all four cases (all p>0.1). The DTI did correlate significantly with evaluations (R=0.35, p<0.001). The DTI did not correlate with any assessment metrics for all WMS who participated in the study, or WMS students who completed all four cases (N=140).

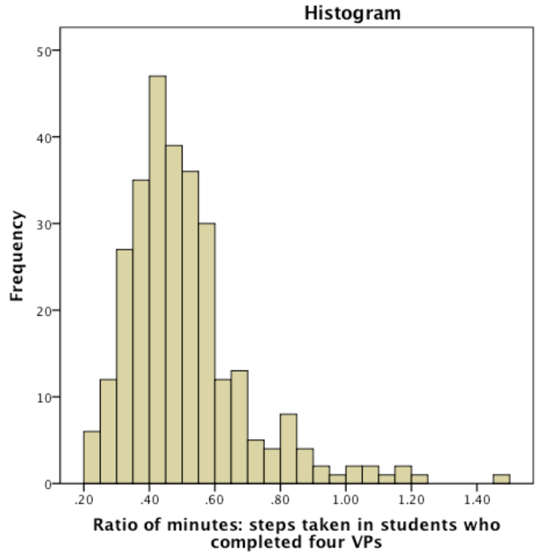
6.2.11.4. Ratio of Time spent per step

During the study it was clear that an additional metric, time spent per step, could be simply calculated from the data. Time spent per step may be a useful reporting metric for future VP work. In this study, for students completing all four VPs, the mean time per step was 0.51 minutes, SD 0.89, n=296. This metric correlated very strongly with the time taken in the case (R=0.97, P<0.00001).

a) Box Plot of minutes:steps ratio in students who completed all four VPs (n=296)



a) Time per step: n=296 students



c) QQ plot of time per step

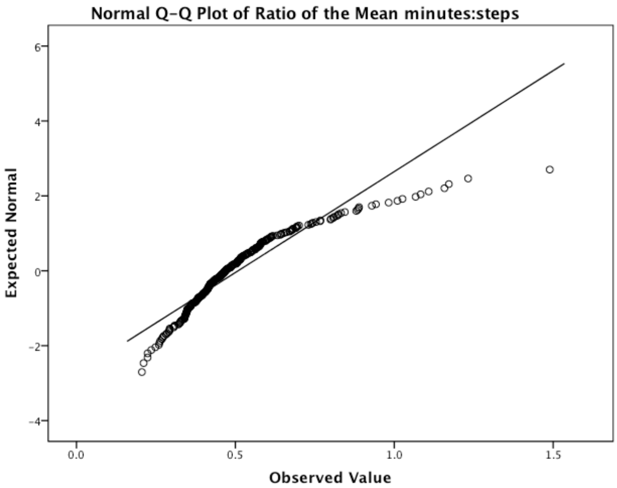


Table 40 Time spent per step: (a) boxplot, (b) histogram, and (c) QQ plot describing the metric in this research study (n=296 students)

6.3. Discussion

This research represents an original contribution to answering how VP design variables influence the effectiveness of VPs in medical undergraduates in the UK. The strengths of this study include the large sample size, multi-centre recruitment, randomisation, adoption of technology standards, open access publication, purpose built research cases, and adoption of established self reported VP evaluation tools. The analysis presents evidence for adopting structured clinical reasoning instruction to improve Bayesian reasoning, and is associated with more positive student evaluations. Branching does not significantly influence performance or evaluation metrics. This research thus has the potential to influence the authoring, adaptation, delivery and use of VPs as both teaching and assessment tools. The only performance metric to meet a 5% improvement set as educationally important in the study protocol (Bateman et al., 2012a) was Bayes reasoning (7% improvement, P<0.001). This improvement was seen where branching was present. Key feature problems, diagnoses, and clinical decisions did not meet this threshold, although some significant differences were noted. This suggests that structured clinical reasoning instruction within cases can significantly improve student's Bayesian reasoning skills, and supports existing teaching approaches (Sedlmeier and Gigerenzer, 2001). This is perhaps a disappointing improvement in Bayesian reasoning, but the intervention was not accompanied with any other supporting information. It does provide evidence that providing students with a frequency grid only (Figure 19) can support Bayesian reasoning.

6.3.1. Primary outcome measures

The primary outcome measures used in this research were performance within the cases, and evaluation of the cases. I used an externally peer reviewed study protocol (Bateman et al., 2012) to plan the research and define important educational effect sizes for performance (a 5% improvement in VP scores) and VP evaluation (a 10% improvement). The research was powered to detect significant differences with 80% power at a (two-sided) 5% significance level. These results show no difference in the adjusted population mean VP scores (ANCOVA) for both branching (absent 8.55, present 8.738, P=0.39) or structured clinical reasoning skills (absent 8.64, present 8.65, P=0.439). The interaction effect between branching and structured clinical reasoning instruction was also non significant. This provides convincing evidence that design variables themselves do not significantly alter global performance within the VP, when adjusting for gender, recruitment centre, and case number. From the analyses of the factorial combinations, there was no significant difference between the four combinations, with the highest score for structured reasoning [present], and branching [present]. I did find a difference in the unadjusted population means in the analysis, with a small difference in performance between structured clinical reasoning instruction (absent vs. present), however this difference was under 2%, lower than the 5% I set as an important effect and educationally relevant.

In terms of student preferences I used the student preference for the best case for learning, and the best case for realism, alongside the EViP evaluation. The most striking finding was that 70% of students preferred a case with structured reasoning instruction (SR) as present as the best case for learning (p<0.001). This was not at the expense of realism, with the students also more likely to list a case with SR present as being more realistic (non-significant). It is important to state that to avoid biasing students by asking if they preferred branching or linear cases, students were asked to nominate the most effective case. For example a student from group 1, choosing case one, would count as: branching, absent; structured clinical reasoning instruction, absent (see Figure 24, p.147). There were no significant differences for either learning or realism for branching (present or absent). The results for the four case designs did show similar scores for the EViP evaluation, with no significant main effect or interaction effects from the adjusted analysis (absent 45.27 present 45.46, P=0.26); branching (absent 45.37, present 45.39, p=0.89); no significant interaction effect (p=0.686). The adjusted (ANOVA) and unadjusted analysis for the EViP evaluation scores failed to show any significant differences between the groups for global or subscale scores (authenticity, professionalism, coaching, and learning). The strongest correlation with EViP evaluation scores was the pre-test self-reported clinical reasoning instruction as measured by the DTI (R=0.3, medium effect size). This suggests that students who rate highly on the metric find the cases a more positive learning experience.

In the ANCOVA analysis when adjusting for fixed factors it was clear that the recruiting centre appeared to significantly influence performance, patterns of use and evaluation. UBMS students scored >8.5% higher than both WMS and KMS students (UBMS, 9.5/15; WMS, 8.3/15; KMS, 8.2/15; see Table 29). These students also spent longer on the VPs (UBMS 32.6 minutes, WMS 27 minutes, KMS 24 minutes), and evaluated the cases with higher EViP scores (UBMS 47/55, KMS 45.1/55, WMS 43.1/55). I believe these results are the first evidence in the literature supporting the impact of curricular integration factors on performance within VPs, and the patterns of use when interacting with them. At the UBMS site, students were introduced to the cases in two timetabled lecture sessions, but free to complete the VPs in their own time, at their own pace outside of a timetabled session. Although the uptake of the cases was lower at the UBMS site, the uptake was arguably surprisingly high for a voluntary learning resource. This uptake in itself at the UBMS site presents evidence of VP case completion as a quality marker.

The conclusion from the primary analysis is that after adjusting for student and case specific factors, the differences between student performance is small. When structured reasoning is present, students are more likely to perform better in some question items (Bayesian reasoning), and prefer SR for learning. For all of the outcomes used, the institution played a more significant role than the case design.

6.3.2. Secondary outcome measures

The data from the secondary outcome measures help explain the relationships between patterns of use and student performance across the cases. Some metrics were not helpful, for example although students took more steps in the cases where structured clinical reasoning was present ($p<0.0001$), this was partly a function of the extra steps used to promote clinical reasoning. These students spent an extra 2.4 minutes per case (branching [absent] 27.5 minutes, branching [present] 29.9 minutes). Students also spent longer on branching cases (0.8 minutes, $P<0.01$). As perhaps would be expected, time spent within a VP did correlate with both the score, and the evaluation of the case (both small effect sizes).

The time spent per case correlated negatively with the scores in the summative end of year written paper (R= -0.22; small- medium effect $p<0.05$). This may reflect that students who naturally spend more time completing clinical reasoning tasks may be subject to increased pressure in a timed examination such as the IPE written, and score lower marks.

The descriptive statistics presented in this research will hopefully improve the design and delivery of VPs, with authors being able to investigate and plan research using these data as a benchmark for performance metrics and patterns of use.

6.3.3. Assessment

I have described a positive correlation between VP scores and performance in summative university written (large effect size, $p<0.001$) and clinical musculoskeletal examination (small-medium effect, $p<0.001$). I have also shown that the quartiles for performance in the VP lead to a convincing differences in summative end of year written papers (quartile 1, 82.47; quartile 2, 89.52; quartile 3, 90.89; quartile 4; 93.15). This practical information in VPs designed predominantly as teaching cases may well be helpful to other institutions planning the use of VPs as assessment tools. This research shows a correlation between VP performance and summative written and clinical assessment.

In 2009, Jonathon Round and colleagues described the use of computer-based scenarios in the USMLE (United States Medical Licencing Examination) stage three as follows (Round et al., 2009):

> *"Despite being a major component of two high stakes exams worldwide, little is known of how they perform in practice. Further evaluation is needed to learn more about this new examination tool." (Round et al., 2009) p.761*

The USMLE stage three requires students to work through a series of questions related to a clinical case. Research in the USMLE stage three has found written and clinical examinations also correlate with this computer-based assessment (Andriole et al., 2005). The use of open-access MedBiquitous compliant VPs in this work presents VPs as accessible and easily distributable resources for formative and summative assessment.

A Best Evidence in Medical Education (BEME) review (Issenberg et al., 2005) highlighted the benefits of physicians receiving feedback on clinical performance as a driver for development. Publishing open access VPs with performance metrics from other students may be a driver for some students to use the cases.

The DTI is important as it is as one of the only validated measurements of clinical reasoning skills (Bordage et al., 1990). One surprising finding is the lack of any correlation between the Diagnostic Thinking Inventory and any marker of performance measured by (1) time spent in the case; (2) key feature problem scores (3) Bayesian reasoning scores; (4) any summative written assessment data; (5) any summative clinical assessment data. Our DTI results 163.5 (SD 17, 95% C.I. 160.5-166.5) are in keeping with those from other research. This includes other UK results (Round et al., 1999, year four students, DTI scores of 160.2 [95% C.I. 158.1-165.8]); US results (158.3 in year 3 students, Bordage et al., 1990); European results (final year Belgian medical students, DTI 168.1, Beullens et al., 2006); and Australian students (DTI 171.1, Groves et al.).

Groves notes that the DTI scores correlate with year of study, but do not predict performance within a cohort (Groves et al., 2002). Beullens et al. (2006) did not find any significant correlation between the DTI and performance from three cohorts. The authors performed a post-hoc analysis but noted that clinical experience was a confounder for DTI, they found a weak correlation only. One research study did find a weak correlation between the DTI and performance in undergraduate Brazilian medical students with both problem solving and summative assessment (Sobral, 1995). As these cases are trying to teach clinical reasoning it is perhaps unsurprising that students who scored high on the DTI also evaluated the cases more positively. These findings clearly link VP performance with summative assessment. However our fifteen-item assessment aligned to validated clinical reasoning assessments does not support the DTI as a measurement of clinical reasoning ability. Whilst unequivocally the DTI increases with training grade, it appears that training grade is a major confounder to its use as a measurement of clinical reasoning ability.

6.3.4. Comparison with instructional design research in VPs

Perhaps the single most influential figure in VP research in the last 10 years has been Professor David Cook (Mayo Clinic, USA). Four of his key papers are quality standards in e-learning healthcare research (Cook, 2005), the first meta-analysis of internet based learning (Cook et al., 2008), the first literature review and proposed research directions for VPs (Cook and Triola, 2009), and the first systematic review of computerised cases for healthcare professionals (Cook et al., 2010). Cook and colleagues perhaps best describe the limitations of existing VP research:

> *"We hope that future researchers can avoid the weaknesses of previous research by designing studies that minimize bias, achieve appropriate power, and avoid confounding." Cook et al., 2010, p.12*

The only study found by Cook to compare student performance was a study of three different simulation cases (Friedman et al., 1991) which is discussed in some detail in 1.2.10 (Discussion of literature not included in the review., p.46). Cook found eleven papers which compared computerised case designs. Most were published before 1990, all used heterogeneous technologies (natural language processing in four papers), many of which are obsolete. Two studies used audiotapes alongside a projector slideshow in dental trainees (Dale et al., 1986, Sandoval et al., 1987). In general the literature was focussed on the adoption of a new technology. In summary there is very little research in this field examining not *if* but *how* VPs should be designed and delivered. I have performed the same literature search in May 2013 as the original review, and identified over 30 additional research publications. This recent work continues the focus on the utility of VPs for delivery in a particular clinical specialty such as psychiatry (Lin et al., 2012), surgery (Yang et al., 2013), dentistry (Cenderberg et al., 2012), and neurology (Johnson et al., 2013). There is virtually no research pubished that provides detailed metrics of cases such as performance scores within the VPs. Perhaps the most significant research findings in VPs in the past decade have been in the area of VP curricular integration, discussed in the next section.

6.3.5. Commentary on curricular integration

Much of the recent literature in VPs has focussed on the impact of different curricular integration strategies (Berman et al., 2009, Edelbring et al., 2012) on student engagement. Berman et al. surveyed over 500 students from six centres, and found that an integration score significantly improved student satisfaction and perceived knowledge gain. Edelbring et al. (2012) looked at different delivery strategies in VPs. This research adds to these findings. It shows that uptake of these computer based cases varies significantly by recruiting centre and VP delivery strategy ($p<0.01$). In the centre with the closest integration, WMS, the VPs were most widely used (WMS, 94.0%; KMS, 83.5%; UBMS, 74.0%). Where integration was at its lowest, the students who volunteered to choose VPs as an additional learning tool spent longer on the cases ($p<0.001$), performed better inside the cases ($p<0.001$), and evaluated them more positively ($p<0.001$). This finding is perhaps counterintuitive, but supports self-selection as a more important determinant to VP evaluation than curricular integration. This work is clearly subject to confounding in that it may be that UBMS students evaluate, perform better, and spend more time on cases for reasons other than curricular integration. The similarities between WMS and KMS, which both involved timetabled VP sessions is, however, striking.

Interestingly at WMS, students who elected not to complete all the four cases, or those that completed no cases did not differ significantly in terms of performance in summative assessments. Independent sample t-tests failed to show any differences between comparisons of the following: 0 cases completed vs. VP completed; 0-1 VP vs. >1 VP; 0-2 VPs vs. 3 or more VPs; all four cases vs. rest of students; all P>0.4. This provides evidence to support that students who do not participate are not simply less academically gifted, but may have other reasons for not completing the VPs.

6.3.6. Commentary on study design authoring and delivery

The factorial study design is a recommended method for research into educational interventions that have multiple design variables (Norman and Eva, 2008), although it has not been widely adopted in medical education research. Thus a PubMed search for ‘medical education’ and ‘factorial’ finds only two papers since 2008 that have adopted this approach in physician and patient education respectively (Zillich et al., 2008, Brooker et al., 2010). I did not find any interaction effect between the two design variables studied. This is unsurprising given the small differences seen between the main effects of each of the independent variables. The research design has doubled the numbers of participants for case comparisons, for example in students completing all four cases (n=296), it enabled comparison of 592 branched cases with 592 linear cases.

I set the size of an important educational effect to be a 5% difference in clinical reasoning scores (0.75/15), and a 10% difference in student EViP evaluation scores. This study exceeded the recruitment targets to detect these changes with 80% power at the (two-sided) 5% level, which required 88 for clinical reasoning and 112 students for self reported evaluation scores (Montgomery et al., 2003). The uptake of the study at the three centres (above 80% actively consenting to participate) and acceptable questionnaire response rates (65.2% from 1184 cases) compare favourably with recent literature: Edelbring et al. (2012) reported a 65% response rate for an end of rotation evaluation.

The case metrics indicate that the case content is appropriate. The intention was to author 30-minute VPs, on average students spent 28 minutes per case. I did not set a prescribed question difficulty in the protocol, although our intention was to write discriminating questions replicating summative assessment. Our students scored around 60% on these cases (mean mark =56.7%). Our figure (57%) is close to the most commonly used pass mark of 60% used by local faculty for medicine clerkships (Kelly et al., 2012), and is appropriately lower than other research where up to 90% of students made the correct diagnoses in clinical reasoning research (Windish et al., 2005).

One interesting reference point is the branching decisions (1 from 3). Our students scored 2.6/4 for clinical decisions. This means for 65% of the times (2.6/4) they would be choosing the correct pathway, and therefore would not be exposed to, or aware of the other branches. For students scoring 4/4 in a branched case, they will have exactly the same experiences as a student scoring 4/4 in a linear case, i.e. they would both take the same pathways. For this reason, both for this study and for future research, the item difficulty of the branching questions is an important metric. It is possible that if the cases were more difficult, or the students were more junior, the branching cases would be more highly evaluated.

6.3.7. Implications for wider VP design and development

The generalisability of these results is supported by the large sample size, multi-centre recruitment, adoption of technical standards, and planned study protocol and statistical analysis. There are numerous factors to consider when authoring a VP as I have shown in the qualitative research. I have identified that when you resource VP development appropriately, and standardise the questions that there is little difference seen between linear and branching cases in the metrics I have chosen. There are significant resource implications for creating VPs with branching structures. The branching cases have four branching points and 81 (3^4) potential routes through them. I have not shown significant differences in performance or evaluation, and only a modest increase in time spent. Further research looking at item difficulty and subject matter may help to determine if branching cases can be more effective in particular scenarios, such as emergency medicine situations or advanced life support.

The use of structured clinical reasoning instruction using evidence-based principles for reasoning is the preferred option for students (Wolpaw et al., 2009; Sedlmeier and Gigerenzer, 2001). Our research does not show any significant differences between case designs on the EViP evaluation, however I did see significant differences between institutions. The publication of open access VP cases showing exactly how the decision-making was conducted has been promoted via the study website (www.go.warwick.ac.uk/msk).

6.3.8. Study Limitations

This study has a number of limitations. Although the majority of students elected to take part (591/719), just over 50% completed all four VP cases.

In terms of a volunteer bias, I have shown from one centre that students who did not participate did not have different scores in summative clinical or written assessments (all p>0.6). This suggests that the study itself was not simply recruiting more academically gifted students, at least in the WMS group. In terms of the 176/572 students who completed a case, but not all four cases (296/572), there was no difference in VP case scores, evaluation scores, or examination results. This again is evidence that these students did not systematically differ from students that remained in the study. Nevertheless this dropout could represent some sort of bias. Equally, although the returned evaluations from our own VPs at 65% equate to other research (Edelbring and colleagues (2012) also had a response rate of 65% in a Swedish cohort) student experience could influence questionnaire response and bias the findings. The study is somewhat complicated in that there is both a primary and secondary analysis for students completing all four VP cases, and for all completed cases. I have tried to present the design and participation in the study in a simple manner, and by adopting a participant flow diagram (Figure 28).

This research uses the VPs as a stand-alone resource to support other teaching modalities. It was presented as replacing some standard faculty presentations in WMS, or as an additional resource and did not represent a major curricular component as in some paediatric education in the United States (Berman et al., 2009). I did not use planned follow up seminars to discuss cases, which has been found to improve evaluation metrics (Edelbring, 2012). In the qualitative work in Section 4 (“

Qualitative Research. Virtual Patients: what works and why? A grounded theory study", p.84) the impact of 'students preconditions' and institutional factors (see Figure 15, 111) appeared to be equally important in shaping their experiences with the cases. The study of VPs at UBMS, in a block where they were an optional extra resource, contrasts with the setting in WMS, where the cases were sat in consecutive teaching sessions. Whilst I argue these two settings reflect a 'real world' context for open access medical resources, these differences both potentially act as confounders.

In terms of VP design although I have adopted and am evaluating VPs developed to open technical standards (Medbiquitous, 2010), I have not evaluated other VP design properties that could fit within this standard, such as the use of video (Bearman, 2003). I have not evaluated computerised case designs that are not MedBiquitous compatible, such as natural language processing (Friedman et al., 1990). Equally other bespoke computer case representations such as virtual three-dimensional environments (Freeman et al., 2001) have not been evaluated. Our cases were approximately 30 minutes long, and I have not studied the optimal length for a VP case as an aim of this research. I also do not know how item difficulty relates to effectiveness or evaluation of VPs, but I do report the item difficulty for all our assessment data, and it is in keeping with commonly used assessment pass marks (Kelly et al., 2012). The VPs I used have been authored in the structured environment of a research project; it is possible that the design properties used would have different implications in VPs authored under more significant time or financial pressures. I have not controlled for factors such as study authorship time, in the sense that I am comparing cases that are branched and linear with the full knowledge that the branching cases had significantly more time resources invested in them.

Although in this thesis I have used the three leading research databases of Medline, PubMed and EMBASE, I have not reviewed the literature from sources not indexed here such as published theses, CINAHL (Cumulative Index of Nursing and Allied Health Literature), or ERIC (Education Resources Information Centre), or conference proceedings. It is probable that individual institutions that have already adopted VP cases may have large databases that contain raw decision data that in magnitude eclipse the case records presented in this study. Nevertheless even within the constructs of standardised case authoring and delivery, I have faced numerous challenges in data collection and management. This study has logged over 100 000 time-stamped individual steps within a VP, and over 25 000 individual clinical decisions on completed cases. One of the challenges in this thesis has been managing and amalgamating individual VP data records.

The case topics formed part of MSK medicine, and focussed on general medical skills including history taking, and interpretation of routine investigation results. Whilst these areas are generalisable to most internal medicine and general practice settings, it is possible that design properties in different clinical areas could affect the conclusions reached in this work. Our students were also year three of a four-year course, and year four of a five-year course. While this is a heterogeneous population of graduate and non-graduate entry students, this research may not be applicable to postgraduate healthcare professionals, more junior students, or students form other allied health professions.

The data analysis of student experience in these cases is based exclusively on quantitative measurements of performance and evaluation. The data analysis on student evaluations also required students to input a unique identifier, their student number. The evaluations themselves were not anonymous and this could potentially bias participants into providing better feedback. The student number was required to track response rates from individual students, cohorts, case designs, and institutions. I have attempted to mitigate bias in that individual case scores and evaluations were not passed back to any faculty in any of the three centres. I did have a number of evaluations (91/1229, 7.4%) that could not be tracked to an individual student: tracking was done by matching a student number inputted by a student against records obtained from each institution. The evaluations also had free text comments. I have read all written comments submitted through the questionnaires, but not subjected them to a formal qualitative analysis. There were however no inappropriate or unprofessional comments made by any students completing any of the 1229 evaluation forms. This research protocol is designed to evaluate only the quantitative results from the EViP questionnaire. I did collect a number of free text responses, which have not been included in the EViP evaluation. A structured qualitative analysis of these free text responses could be used to further inform VP design principles.

There were other examples of missing or unusable records data throughout the study that reflected technical or logistical problems with the research design. Whilst a limitation, where data was missing it was generally from a small proportion of cases. In all cases, missing data was never above 10% of the total number of metrics collected. There was missing performance data for <2% for the total cases completed (1773 complete cases). In addition for WMS, <4% of end of year written and clinical assessments, and <8% of the end of block assessments, and <10% of EViP evaluations could not be tracked to an individual. A tangible example of the unpredictability of external factors was the fact institutional IT policies can impact on the learner experience (Bateman and Millett, 2009). One hospital trust blocked access to VPs from certain computers as they were incorrectly identified as a 'retail estate agent' by automated Internet security software.

This study design here evaluates student performance within cases, and their evaluation of cases. I have not assessed learning in a pre- and post-test assessment. This was a conscious decision reflecting that VPs are effective as web based educational interventions (Cook et al., 2010). The factorial study design was intended to expose all students to the four potential VP designs. The only metric recorded as a post intervention measure was the DTI, which showed a non-significant small improvement over 8 weeks (mean improvement =0.57, p=0.71). The DTI in this cohort does not correlate with any direct or indirect quantitative measurement of clinical reasoning during this study. It should be noted that the original DTI was a paper assessment and that I delivered a web based version of the questionnaire. I had anticipated a poor response rate for the DTI for two reasons.

Firstly the students did not achieve any obvious benefit from completing the DTI, and secondly, as a 20-minute, 41-item questionnaire, asking students to complete two such forms voluntarily was likely to be challenging. I did not offer the DTI to students from WMS or UBMS, as I felt the 20 minute pre-test evaluation could potentially negatively affect enrolment to the study. For similar reasons at WMS, when the VPs were being introduced to the first cohort of students, I did not use the DTI to avoid an initial negative reaction to the cases. In total collecting 134 complete DTI reports from 166 WMS students (89.3%) was a good response, with paired pre-post test data available for 83 students (55%).

Problems with questionnaire validation could equally be levelled at the EViP questionnaire developed by Huwendiek and de Leng (2010), which to my knowledge have not been formally validated. I do present data from this research supporting the good 'internal consistency' of the questionnaire (see footnote to Table 18, p.169). This refers to the extent to which the items relate to a similar construct. This does not constitute clear validity evidence, ideally I would have a range of VPs of different lengths, case structures, and authors to compare (DeVellis, 2003, Cook and Beckman, 2006). Given the existence of different self-reported evaluations to assess VPs in the recent literature, (Berman et al., 2009, Huwendiek and de Leng, 2010., Edelbring et al., 2012) a work to validate a VP assessment would be a useful future research project.

Overall the study protocol provided some flexibility to allow students to participate and complete cases in different settings, and thus facilitate recruitment from different universities. Case and evaluation completion rates reflect this, and I believe the transparent data reporting, including the missing data at the various stages adds to the strength of the findings.

6.3.9. Summary

The key findings from the study are shown to the right in Figure 37. In particular students preferred structural reasoning for learning when choosing an individual case best for learning (see Table 25, p.185). I have also shown that comparisons between global case designs do not make a significant difference in overall global performance or evaluation scores (Table 29, p.192; Table 31 p.195).

What was already known on this topic
• VPs are effective learning tools • There are numerous VP design typologies • Recent technology standards represent a paradigm shift for VP delivery
What this research adds
• Design typologies produced non-significant changes to VP performance and EViP evaluation metrics. • Students preferred structured clinical reasoning instruction for learning • VPs scores correlate significantly with summative clinical and written assessments • Curricular integration strategies seem to significantly influence performance, use patterns and evaluation of VPs • Pre-study DTI scores did not predict performance in VPs, or any summative assessments. • This research supports the use of open access VPs for MSK education. Further research may explore case length, difficulty and integration.

Figure 37 Key research findings

This research is an example of a factorial design classically used for drug therapy research to compare the interaction between treatments (Montgomery et al., 2003). I have applied a design explicitly called for by education researchers (Norman and Eva, 2008). In this research I found no significant interaction effects between the design variables in the analysis.

VP design did impact on patterns of use, particularly for 'structured clinical reasoning instruction' (Table 26, p.187): students spent more time on these cases, taking more steps. In this student population VP scores correlate significantly with summative clinical and written assessments (see Table 36, p.202). To this extent the research has fulfilled the original stated protocol, successfully recruiting the desired number of students to answer the original research question. These research findings present the quantitative impact of different VP designs. The cases that have been developed are available on open access basis.

Section 7. Conclusions and proposals for future work

Both the qualitative and quantitative research findings presented support future VP design and delivery of virtual patients in undergraduate education. This concluding section debates the originality, relevance and the impact of this research in virtual patients. This section begins with a summary of the individual components of the research, and moves on to discuss the practical messages for VP authors. The findings from each of the two separate research studies are then contrasted against each other to help provide an overview of the research and provide additional validity evidence for the quantitative model, drawing the two separate research components together. This section closes with a description of the potential next steps for this research.

7.1. Summary of research findings

This research has achieved the original goals outlined in a research proposal to the extent that I have conducted and completed a qualitative and quantitative research project investigating the effectiveness of VPs. The qualitative research (section 4) is a single centre grounded theory study into MSK education using virtual patients. It is original in three principle ways. Firstly, it is the first grounded theory study into virtual patients. Secondly, it is also the first research study investigating VP designs using VPs explicitly created for researching individual design properties. Thirdly the VP cases were submitted to the publisher as ‘supplementary material’ to the Open Access paper (Bateman et al., 2013), available to readers to download. To my knowledge this is the first research study of VPs to publish the cases as Open Access resources in this way. The resulting model (Figure 15, page 111) describes how students experience VPs in three layers. The inner layer describes the influence of the students 'pre-conditions' before sitting a VP. The middle layer describes encoded and constructed activity. Encoded activity contains what can be authored into a case. Constructed activity relates to how, when and where an institution uses a VP. The outer layer describes cognitive and behavioural change. This model provides an original contribution to knowledge about VPs for both authors and institutions. The model can help faculty plan authoring, development, repurposing, delivery, and curricular integration of VPs.

Section 5 discussed the authoring of cases and the development of a study protocol (Bateman et al., 2012) that is included in the appendix 8. This process was critical to addressing the challenges laid out by Cook and colleagues (2010) in a systematic review of VP cases. I have addressed the challenges in defining typologies of VP cases (Huwendiek et al., 2009a) and produced a standardised reporting template.

What was already known on this topic
• VPs are best placed to teach clinical reasoning skills • We do not know how they should best be designed
What this research adds
• The model described how VP design should be considered in as part of one component of how a student learns from a VP • All 'branching' and 'structured clinical reasoning instruction' (SR) factorial design combinations were positively evaluated by students, with no differences in global performance or case evaluations • SR was associated with improvements in Bayesian reasoning, and was the preferred option for 'learning' when asked to select a single design • Authors should consider routinely adopting SR , but not branching • VPs may have a role in assessment and appraisal of students • Further research may explore: VP integration; open educational resources; and organisational change

Figure 38 Summary of research

Finally in Section 6 I have presented the methods and results from an original multi-centre randomised research study investigating the effectiveness of two independent VP designs. This original research is the first study of its type using the technical standards that define VPs. It is one of the largest research studies ever conducted into computerised case design in the health professions. My conclusions, based on this research, are that the optimal design of VPs for undergraduate education should include structured clinical reasoning instruction, but that branching should not be routinely used. Structured clinical reasoning instruction is preferred as the best design for learning by 70% of students who completed all four cases (Table 25, p.185) and did not influence the perceived realism of the case. I conclude that different branching pathways should not be used routinely in undergraduate education because despite the significant resource allocation to author such cases there is no improvement in performance (Table 29, p.192), self-reported evaluation (Table 31, p.195), or patterns of use (Table 27, p.189). The appeddices (p. 262) gives a representation of how difficult the cases are to construct. I estimate the total case authoring time doubles for two branches and trebles for three branches.

I found significant correlations between summative assessment results and student performance within the case at one centre (Table 36, p.202). Interestingly I have found no correlation between measured performance in cases and summative assessments with self-reported measurements of clinical reasoning ability (Table 38, p.205).

These VPs were authored using a structured system that includes a peer review process. When authoring standards, editing, peer review, and prior authoring experience are factored in, the impact of these design principles shows small differences between the VP typologies for VP performance scores, and evaluation. I would add a cautionary note about the findings from branching cases. Their use in different settings, case topics, and with different item difficulty, and their use in adaptive settings, may all positively or negatively influence their effectiveness. The research reinforces that VP authors should consider the difficulty of branching nodes, in that if the item difficulty is low, few students will experience the branching component. My conclusion is that the relatively large difference between cohorts is most likely to reflects differences relating to curricular integration of the cases in the three centres, rather than differences in performance. This would be a subject of further research. I have also described how VP performance scores predict summative assessment results for clinical and written assessments. Whilst all of this research is in MSK medicine, a sphere that includes rheumatology and orthopaedics, I believe these results are likely to be generalisable to other medical and surgical specialities at undergraduate level.

7.2. How this research informs VP authoring

The components from these two original research studies provide new knowledge about the authorship and delivery of VPs that may be relevant to authors, faculty and institutions. This qualitative model is supported by the quantitative research presented identifying differences in student patterns of use, performance and evaluation between centres. This research has been as the subject of an editorial in the journal 'Medical Education' titled "Research into the use of virtual patients is moving forward by zooming out" (Edelbring, 2013), and a research digest article in the journal 'The Clinical Teacher' titled: "Virtual patient design, what works and why" (2013). The adoption of the structured reasoning processes 'SNAPPS' (Wolpaw et al., 2010) alongside a frequency grid (Sedlmeier and Gigerenzer, 2001) is practical, and appears to influence student experiences and performance.

To some extent these research findings challenge the role of the DTI as a measurement of clinical reasoning performance. This confirms the findings from European and Australian research studies discussed in Section 6.2.11.3 (p. 205), however other research conflicts with these conclusions (Sobral, 1995). The present study compares the DTI explicitly against clinical reasoning scenarios. Not only is there no correlation with VP case scores, there is no correlation with written and clinical summative assessment in MSK or in general medicine. I also present some of the first quantitative research into the use of branching cases. This original work suggest that educators should consider carefully the goals and resource implications of authoring branching cases, given the metrics I have recorded in these structured VPs.

This research will ideally be published in open access form in an established journal to help disseminate the research findings. The use of open access cases by different institutions is likely to be facilitated by the open availability of the cases on the Internet. The four VP topics are available at www.go.warwick.ac.uk/msk. The on-going use of these cases outside of the research environment is measurable, and the qualitative work can potentially inform its development. I present a series of examples of how the research is being disseminated in **Error! Reference source not found.** (p.**Error! Bookmark not defined.**), including student videos, social media engagement, and a website.

7.3. Lessons from comparing and contrasting the findings from the qualitative and quantitative research

The qualitative and quantitative research sections in this thesis were conducted on students from three centres: WMS, UBMS and KMS. This subsection reflects on the findings from these two separate studies, comparing and contrasting the results from the model produced (Figure 15, p.111) and the quantitative research. To recap, the qualitative model describing how a student learns from a VP (figure 15, p.111) has three layers. The inner layer describes individual learners that will go on to use the VP ('student preconditions'). The middle layer describes the VP design itself ('encoded activity'), and how VPs are delivered to students ('constructed activity'). The outer layer represents cognitive and behavioural change in students after interacting with a VP. It is possible to apply the model to the students in the quantitative study, by exploring the interaction between each of these layers between the different ‘constructed activity’ at Warwick, Birmingham and Keele. For example student preconditions can be informed from student, constructed activity from the institution, and student-VP interaction from data collected on patterns of use, performance, and completion. Throughout the study the ‘encoded activity’, or VP case design, was controlled for whereas the ‘student preconditions’ and ‘constructed activity’ changed across the three recruiting sites (see Figure 27, p.153). For example, at WMS, the students were all graduate entry medical students. The KMS and UBMS students were predominantly undergraduate entry students with less than 20% of the sampled cohort graduate entry at both sites (see table 12, p. 145; table 15, p. 162). At WMS and KMS students had timetabled teaching sessions, whereas at UBMS the students completed the cases in their own time. This allows for a number of comparisons to be made.

7.3.1. Constructed activity is associated with predicted differences in VP use and student satisfaction

The qualitative model predicts that as the ‘constructed activity’ varied across the sites it would impact on the ‘student-VP interaction’. The qualitative research suggested these changes would be measurable (Figure 13, p.105). As predicted by the model, there were significant, meaningful differences between the universities in recruitment (table 15, p.162), performance (Figure 35, p.199; table 35, p.200), and evaluation (table 35, p.200). The controlled study design means this is most likely explained by the ‘student preconditions’ and ‘constructed activity’. The proportion of students that went on to complete all four cases was significantly different across the three schools, ranging from completion rates of 62% for WMS, 42% for UBMS, and 40% at KMS (Table 15, p.162). A decision to complete further VPs and engage with them as a form of learning was a component of 'behavioural change' that I coded and defined as "*Personal 'Buy In'-Extent to which VP design influences current and future participation of VP cases*" (Table 9, p. 109). These case completion data support that completion rates are influenced by constructed activity.

At UBMS, students also returned significantly higher self-reported evaluation scores for VPs, spent significantly longer on the VPs, scored significantly higher marks on cases than at the other two universities (Figure 35, p.199). As highlighted earlier, at UBMS the 'pedagogic use' was focussed on the cases as an additional optional learning resource. Surprisingly, the UBMS students had non-inferior (slightly higher) study completion rates when compared to KMS, despite not having a dedicated teaching session. It appears therefore that the UBMS students who completed the cases valued this flexible 'constructed activity' for VPs, the ability to personally decide when and where to learn from the VPs. The VP data logs support this, with a significant proportion of cases at UBMS being completed at the weekend (19.8%). In summary the constructed activity appears to influence uptake, experience and learning experiences from the VPs as predicted by the model.

7.3.2. Lessons looking globally at case metrics

As predicted by the model, there were no gender differences for VP use, engagement or performance (Figure 29, p. 165). Students that spend longer on the cases performed better in them, and evaluated them in a more positive light (both $P<0.001$, see table 38, p.205). This is consistent with the model that described that ‘skipping’ of case content (Table 8, p.103) reflected a lack of engagement (Figure 13, p.105), and would inevitably predict poorer performance because of missing data and lower case evaluation. It is not clear if the ‘mind wandering’ or task unrelated thoughts seen in these VPs translate to mind wandering that can be harmful in real clinical practice (Further analysis of individual steps in the cases could help support this hypothesis further.

7.3.3. The model, assessment metrics, and limitations

The qualitative research did not identify a student’s academic ability as an important factor for case completion or engagement. This is supported by the quantitative results showing case completion rates at WMS were not influenced by the score in the VP, and did not predict future performance in summative assessment scores (Table 23, p.176). This includes the performance in domain specific exams (MSK written assessment, MSK OSCE), and in general written and clinical summative assessments. Conversely for those students who did complete all four cases, the performance in the VPs did predict written and clinical performance (Figure 36, p.203). This supports two conclusions. Firstly, for the students that do not engage with VPs the explanation is not baseline academic ability. Secondly in the students who complete a sequence of VPs, VP performance correlates with predict summative clinical and written examinations. These two conclusions potentially inform future research directions for how to deliver and support VPs in teaching and assessment. Unfortunately this research cannot provide direct evidence to support cognitive and behavioural change predicted by the model as a result of completing the cases.

Despite the limitations and potential sources of bias that I have outlined in both qualitative (4.5.4, p.126) and quantitative research (6.3.8, p.222), I have demonstrated that the multi-centre findings do support to some extent the validity of the qualitative model. It remains to be seen if, and how approaches can be formulated for identifying lessons from individuals and institutions to support the delivery of VPs. Undoubtedly the use of performance metrics in demonstrated in this work may help other researchers plan future research studies.

7.4. Next steps

I would like to continue with research into VPs, and I would consider that there are a number of unanswered questions in VP research. I would argue there are two important further questions: addressing an improvement in patient care; and addressing uptake of OERs (Open Educational Resources).

7.4.1. Question 1. Can the VPs produced using the lessons from this research demonstrate tangible benefits to patients?

The BEME Collaboration define four levels of evidence for education research using the Kirkpatrick Hierarchy (1967), available online (BEME Collaboration, 2005). The BEME evidence group describe the hierarchy as four levels (see Table 41, p.242).

"Level 1 *Participation - covers learners' views on the learning experience, its organization, presentation, content, teaching methods, and aspects of the instructional organization, materials, quality of instruction* ***Level 2a*** *Modification of attitudes/perceptions - outcomes here relate to changes in the reciprocal attitudes or perceptions between participant groups toward intervention/simulation* ***Level 2b*** *Modification of knowledge/skills - for knowledge, this relates to the acquisition of concepts, procedures and principles; for skills this relates to the acquisition of thinking/problem-solving, psychomotor and social skills* ***Level 3*** *Behavioural change - documents the transfer of learning to the workplace or willingness of learners to apply new knowledge & skills.* ***Level 4a***

Change in organizational practice - wider changes in the organizational delivery of care, attributable to an educational program ***Level 4b*** *Benefits to patient / clients - any improvement in the health & well being of patients/clients as a direct result of an educational program."* Reproduced from: BEME Collaboration, 2005, Accessed 03.04.2012

Table 41 The BEME collaboration hierarchy of evidence

The present research was not designed to address the higher levels of the Kirkpatrick Hierarchy, I made an assumption using the evidence from a meta-analysis that suggests internet-based learning is effective (Cook et al., 2008). A logical next step for the research would be to use the model, and quantitative research evidence to explore if VPs designed using these principles can produce Kirkpatrick level 3 or 4 outcomes. Kirkpatrick Level 3 and 4 research in medical education is rare, one review found research investigating these levels occurring in <10% medical education research articles (Dornan et al., 2006), before any quality criteria were applied. Kirkpatrick change is dependent on the research having appropriate quality criteria (Yardley and Dornan, 2012). I have presented an example study protocol to research this in Figure 39.

Problem statement. Inflammatory low back pain (ILBP) is a cause of <5% of low back pain (LBP), a common presenting feature to primary care. ILBP is frequently missed, leading to the late diagnosis of ankylosing spondylitis, or AS (Sieper and Rudwaleit, 2005). The blood test HLA B27 can help stratify the risk of patients with particular symptoms of AS. VPs can teach clinical reasoning skills, but Kirkpatrick levels 3 and 4 evidence to support their use is limited. Recent VP research has produced both a model for designing and delivering VPs, and case authoring recommendations.
Research Question. Can VPs used to teach ILBP assessment and clinical reasoning using blood tests like HLA B27 improve testing, referral and diagnosis of AS in a regional cohort.
Hypothesis. An ILBP VP could improve use of HLA B27, early referral and diagnosis rates for AS.
Research Design. Randomised prospective two-group experimental study of low back pain assessment by General Practitioners in the West Midlands, facilitated by either a Virtual Patient (VP) or standard education resources.
Intervention. Linear Virtual patient (15 minutes) using structured clinical reasoning instruction, and Bayesian reasoning frequency grid to explore ILBP.
Control. Standardised educational leaflet containing clinical information in the VP, to be emailed and posted to practitioners.
Population. General practitioners in the local Clinical Commissioning Group (CCG).
Randomisation. Computerised randomisation of all participants.
Primary outcome measures. (1) HLA B27 tests requests; (2) referral rates to rheumatology for back pain; and (3) new AS diagnoses.
Secondary outcome measures. (1) VP patterns of use, (2) EViP VP evaluation
Sample size and power calculation. To be confirmed, anticipate recruiting 50 general practitioners to the study, over a period of 12 months. Formal power calculation to be performed.
Statistics. Appropriate parametric and non-parametric statistics, including independent samples T test, Chi square. A statistical analysis plan will be provided in the protocol.
Data collection and quality. Led centrally by the primary investigators.
Timetable and Budget. 1 year, applications for funding in process
Protocol. To be completed

Figure 39 Example preliminary study protocol to Kirkpatrick levels 3 and 4 change

A criticism of this research study would be the logistical challenges, cost and risking a return to the historical error of researching “if” e-learning works (Cook et al., 2008). Nevertheless if we are to educate and revalidate doctors, there must be evidence for the most effective learning tool, and it is possible that research of this nature could persuade regulatory bodies of the importance of cases both in MSK medicine and other specialties.

7.4.2. Question 2. How can we best facilitate and evaluate national adoption of open access VP resources?

VPs are effective teaching clinical reasoning skills, but the best way to promote widespread adoption from healthcare professionals is not clear. Research into different cohorts of students completing large numbers of VP cases may well provide evidence to help inform VP design, delivery, and curricular development in different schools. The evaluation tools, whist not perfect, have delivered usable measurable data for three centres, which should be suitable for escalating to a national level. There is an appetite for open-access publication and OERs. Open publication is now mandated by the UKs leading charitable research foundations (Hawkes, 2012), OERs are supported by international bodies (UNESCO, 2011). National research studies are perhaps supported and funded by an independent respected national body such as Jisc® (formerly JISC, the Joint Information Systems Committee) or a large research charity, such as Arthritis Research UK, who have funded this research.

Open publication itself is challenging. Jisc® has already part funded research to provide practical assistance for OER publishing in public health (Hemingway et al., 2011). The authors produced a framework for people interested in creating OERs, stating:

> *"We hope that this will provide practical assistance and encouragement for the academic public health community to create and share OER."*
>
> *(Hemingway et al., 2011), p.42*

Despite receiving research funding from a publicly funded body, the article by Hemingway et al. (2011) is published behind a paywall (requires a paid subscription), and the framework proposed is copyrighted to the publisher, not the authors. This contemporary example highlights the challenges faced by OER. Within the VP environment, although there are open-access VP development tools available (Begg, 2010) they do not provide the functionality of the commercial players that were used to author cases during this study. The changing landscape is understood even more clearly in that the original VP player used for this research was the software

‘VpSim’, created in house at the University of Pittsburgh. An independent subsidiary company (DecisionSim LLC) went on to develop further iterations of the player. It could be argued that these developments represent both opportunities and threats to OER development and research.

The evaluation of VP use across local and national cohorts could potentially act as a quality marker for education. Numerous challenges face MSK education in the UK alone (Dacre and Fox, 2000, Adebajo et al., 2005) including the rising medical student intake (Badcock et al., 2006).

Any national resource would require hosting, either at a single webpage, or via another institution or organisation. I have made the cases available on University hosted web pages adopting a “user centred design” (Kinzie et al., 2002). They are simple to access through the link www.go.warwick.ac.uk/msk. This local hosting provides flexibility in delivery. Curated peer reviewed OERs are available from sources such as MedEdPORTAL (American Association of Medical Collages, 2012), or through the Jisc® (formerly JISC) website. The advantages of making resources available for repositories are that this allows visibility, indexing and searchability for other interested parties. OERs may be particularly relevant to developing countries (Brower, 2010) , in particular for e-learning (Bhan, 2005), and VPs (Bediang et al., 2012). In terms of other approaches to publicising and advertising cases to faculty and students, social media is a logical choice for a web based tool (Scott, 2013). The first ever systematic review of social media in medical education (Cheston et al., 2013) suggests it may promote learner engagement, feedback, and collaboration.

In summary, these two research studies present the first steps on a journey of understanding to how best design, deliver, integrate, standardise, and evaluate the realistic representation of clinical cases online: virtual patients.

Section 8. References

Adebajo, A., Windsor, K., Hassell, A. & Dacre, J. 2005. Undergraduate education in rheumatology. *Rheumatology,* 44**,** 1202-3; author reply 1203.

American Association of Medical Collages. 2012. *Mededportal* [Online]. Available: https://www.mededportal.org/ [Accessed 05.01 2012].

Andriole, D. A., Jeffe, D. B., Hageman, H. L. & Whelan, A. J. 2005. What predicts usmle step 3 performance? *Academic Medicine : Journal of the Association of American Medical Colleges,* 80**,** S21-4.

Arthritis Research UK. 2010. *Development of a web based tool for teaching musculoskeletal medicine skills to medical students-the 'virtual patient'* [Online]. Available: http://www.arthritisresearchuk.org/research/grant-tracker-items/2010/development-of-a-web-based-tool-for-teaching-musculoskeletal-medicine-skills-to-medical-students--the-virtual-patient.aspx [Accessed 5th January 2011].

Badcock, L. J., Raj, N., Gadsby, K. & Deighton, C. M. 2006. Meeting the needs of increasing numbers of medical students--a best practise approach. *Rheumatology,* 45**,** 799-803.

Banning, M. 2008. The think aloud approach as an educational tool to develop and assess clinical reasoning in undergraduate students. *Nurse Education today,* 28**,** 8-14.

Bannister, S. L., Hilliard, R. I., Regehr, G. & Lingard, L. 2003. Technical skills in paediatrics: A qualitative study of acquisition, attitudes and assumptions in the neonatal intensive care unit. *Medical Education,* 37**,** 1082-90.

Barrows, H. S. & Feltovich, P. J. 1987. The clinical reasoning process. *Medical Education,* 21**,** 86-91.

Bateman, J., Allen, M., Samani, D., Kidd, J. & Davies, D. 2013. Virtual patient design: Exploring what works and why. A grounded theory study. *Medical Education,* 47**,** 595-606.

Bateman, J., Allen, M. E., Kidd, J., Parsons, N. & Davies, D. 2012a. Virtual patients design and its effect on clinical reasoning and student experience: A protocol for a randomised factorial multi-centre study. *BMC Medical Education,* 12**,** 62.

Bateman, J. & Davies, D. 2011. Virtual patients: Are we in a new era? *Academic Medicine : Journal of the Association of American Medical Colleges,* 86**,** 151; author reply 151.

Bateman, J., Davies, D. & Allen, M. 2012b. Mind wandering has an impact in electronic teaching cases. *Medical Education,* 46**,** 235.

Bateman, J., Francis, R. & Thistlethwaite, J. 2011. Medical student burnout and professionalism. *JAMA : the Journal of the American Medical Association,* 305**,** 37-8; author reply 38.

Bateman, J. & Millett, A. 2009. Educational institutions' information technology policies have the potential to impede e-learning research. *Medical Teacher,* 31**,** 774.

Bateman J., on Behalf of the European Virtual Patients Project Team. 2010. *Open access VP cases: Mr Cooper, Wadih Hakim, Maria Gallon, Sarah Gillespie*

[Online]. Available: http://www.virtualpatients.eu.org/referatory [Accessed 12th August 2010].

Bearman, M. 2003. Is virtual the same as real? Medical students' experiences of a virtual patient. *Academic Medicine : journal of the Association of American Medical Colleges,* 78, 538-45.

Bearman, M., Cesnik, B. & Liddell, M. 2001. Random comparison of 'virtual patient' models in the context of teaching clinical communication skills. *Medical Education,* 35, 824-32.

Bediang, G., Raetzo, M. A. & Geissbuhler, A. 2012. Virtual patient simulation: A comparison of two approaches for capacity building in sub-saharan africa. *Studies in health technology and informatics,* 180, 978-82.

Begg, M. 2010. Virtual patients: Practical advice for clinical authors using labyrinth. *The clinical teacher,* 7, 202-5.

Beme Collaboration. 2005. *Beme coding sheet: Features and uses of high-fidelity medical simulations that lead to effective learning: A beme systematic review* [Online]. Available: [Accessed 03.04.2012 2012].

Bergman, E., De Fleijter, J., Frambach, J., Cgodefrooij, M., Slootwek, I., Stalmeijer, R. & Cvan Der Zwet, J. 2012. AM last page: A guide to research paradigms relevant to medical education. *Academic Medicine : Journal of the Association of American Medical Colleges,* 87, 1.

Berman, N., Fall, L. H., Smith, S., Levine, D. A., Maloney, C. G., Potts, M., Siegel, B. & Foster-Johnson, L. 2009. Integration strategies for using virtual patients in clinical clerkships. *Academic Medicine : Journal of the Association of American Medical Colleges,* 84, 942-9.

Berman, N. B., Fall, L. H., Chessman, A. W., Dell, M. R., Lang, V. J., Leong, S. L., Nixon, L. J. & Smith, S. 2011. A collaborative model for developing and maintaining virtual patients for medical education. *Medical Teacher,* 33, 319-24.

Berman, N. B., Fall, L. H., Maloney, C. G. & Levine, D. A. 2008. Computer-assisted instruction in clinical education: A roadmap to increasing cai implementation. *Advances in Health Sciences Education : Theory and Practice,* 13, 373-83.

Beullens, J., Struyf, E. & Van Damme, B. 2006. Diagnostic ability in relation to clinical seminars and extended-matching questions examinations. *Medical Education,* 40, 1173-9.

Bhan, A. 2005. Importance of open access for clinicians and researchers in developing countries. *CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne,* 173, 129.

Bian, Z. X. & Shang, H. C. 2011. Consort 2010 statement: Updated guidelines for reporting parallel group randomized trials. *Annals of internal medicine,* 154, 290-1; author reply 291-2.

Birks, M. & Mills, J. 2011. *Grounded theory : A practical guide,* Los Angeles, Calif. ; London, SAGE.

Bitzer, M. 1966. Clinical nursing instruction via the plato simulated laboratory. *Nursing research,* 15, 144-50.

Bondy, K. N., Koenigseder, L. A., Ishee, J. H. & Williams, B. G. 2001. Psychometric properties of the california critical thinking tests. *Journal of nursing measurement,* 9, 309-28.

Bordage, G., Grant, J. & Marsden, P. 1990. Quantitative assessment of diagnostic ability. *Medical Education,* 24**,** 413-25.
Botezatu, M., Hult, H. & Fors, U. G. 2010a. Virtual patient simulation: What do students make of it? A focus group study. *BMC medical education,* 10**,** 91.
Botezatu, M., Hult, H., Tessma, M. K. & Fors, U. 2010b. Virtual patient simulation: Knowledge gain or knowledge loss? *Medical Teacher,* 32**,** 562-8.
Botezatu, M., Hult, H., Tessma, M. K. & Fors, U. G. 2010c. Virtual patient simulation for learning and assessment: Superior results in comparison with regular course exams. *Medical Teacher,* 32**,** 845-50.
Bowen, G. A. 2006. Grounded theory and sensitizing concepts. *International Journal for Qualitative Methods,* 5**,** 1-9.
Bridgemohan, C. F., Levy, S., Veluz, A. K. & Knight, J. R. 2005. Teaching paediatric residents about learning disorders: Use of standardised case discussion versus multimedia computer tutorial. *Medical Education,* 39**,** 797-806.
Brooker, S., Okello, G., Njagi, K., Dubeck, M. M., Halliday, K. E., Inyega, H. & Jukes, M. C. 2010. Improving educational achievement and anaemia of school children: Design of a cluster randomised trial of school-based malaria prevention and enhanced literacy instruction in kenya. *Trials,* 11**,** 93.
Brooks, L. R., Norman, G. R. & Allen, S. W. 1991. Role of specific similarity in a medical diagnostic task. *Journal of experimental psychology. General,* 120**,** 278-87.
Brower, S. M. 2010. Medical education and information literacy in the era of open access. *Medical reference services quarterly,* 29**,** 85-91.
Bryant, A. & Charmaz, K. 2007. *The SAGE handbook of grounded theory,* London, SAGE.
Bunnis, S. & Kelly, R. 2010. Research paradigms in medical education research. *Medical Education,* 44**,** 8.
Callahan, C. A., Hojat, M. & Gonnella, J. S. 2007. Volunteer bias in medical education research: An empirical study of over three decades of longitudinal data. *Medical Education,* 41**,** 746-53.
Campbell, G., Miller, A. & Balasubramaniam, C. 2009. The role of intellectual property in creating, sharing and repurposing virtual patients. *Medical Teacher,* 31**,** 709-12.
Centre for Reviews and Dissemination 2009. *Systematic reviews: Crd's guidance for undertaking reviews in health care. ,* York, University of York.
Cenderberg, R. A., Bentley, D. A., Halpin, R., Valenza, J.A. 2012. Use of virtual patients in dental education: a survey of US and Canadian dental schools. *J Dent Educ.* 76(10):1358-64.
Charmaz, K. 2003. Grounded theory: Objectivist and constructivist methods. *In:* Denzin, N. K. & Lincoln, Y. S. (eds.) *Strategies of qualitative inquiry.* 2nd ed. Thousand Oaks, CA: SAGE.
Charmaz, K. 2004. Premises, principles, and practices in qualitative research: Revisiting the foundations. *Qualitative health research,* 14**,** 976-93.
Charmaz, K. 2006. *Constructing grounded theory : A practical guide through qualitative analysis,* London ; Thousand Oaks, SAGE.
Cheston, C. C., Flickinger, T. E. & Chisolm, M. S. 2013. Social media use in medical education: A systematic review. *Academic Medicine : Journal of the Association of American Medical Colleges*.

Cilliers, F. J., Schuwirth, L. W. & Van Der Vleuten, C. P. 2012. Modelling the pre-assessment learning effects of assessment: Evidence in the validity chain. *Medical Education,* 46**,** 1087-98.

Clark, R. C. & Mayer, R. E. 2008. *E-learning and the science of instruction : Proven guidelines for consumers and designers of multimedia learning,* San Francisco, Pfeiffer.

Clinical Teacher, 2013. Virtual patient design, what works and why. *The Clinical Teacher,* 10**,** 274.

Cohen, J. 1992. A power primer. *Psychological Bulletin,* 112**,** 155-9.

Connick, R. M., Connick, P., Klotsas, A. E., Tsagkaraki, P. A. & Gkrania-Klotsas, E. 2009. Procedural confidence in hospital based practitioners: Implications for the training and practice of doctors at all grades. *BMC Medical Education,* 9**,** 2.

Conole, G., Oliver, M. & Taylor & Francis. 2006. Contemporary perspectives in e-learning research themes, methods, and impact on practice. *The open and flexible learning series.* Abingdon, Oxon ; New York: Routledge,.

Cook, D. A. 2005a. Learning and cognitive styles in web-based learning: Theory, evidence, and application. *Academic Medicine : Journal of the Association of American Medical Colleges,* 80**,** 266-78.

Cook, D. A. 2005b. The research we still are not doing: An agenda for the study of computer-based learning. *Academic Medicine : Journal of the Association of American Medical Colleges,* 80**,** 541-8.

Cook, D. A. & Beckman, T. J. 2006. Current concepts in validity and reliability for psychometric instruments: Theory and application. *The American journal of medicine,* 119**,** 166 e7-16.

Cook, D. A., Erwin, P. J. & Triola, M. M. 2010. Computerized virtual patients in health professions education: A systematic review and meta-analysis. *Academic Medicine : Journal of the Association of American Medical Colleges,* 85**,** 1589-602.

Cook, D. A., Levinson, A. J., Garside, S., Dupras, D. M., Erwin, P. J. & Montori, V. M. 2008. Internet-based learning in the health professions: A meta-analysis. *JAMA : the journal of the American Medical Association,* 300**,** 1181-96.

Cook, D. A. & Triola, M. M. 2009. Virtual patients: A critical literature review and proposed next steps. *Medical Education,* 43**,** 303-11.

Corbin, J. M. & Strauss, A. L. 2008. *Basics of qualitative research : Techniques and procedures for developing grounded theory,* London, SAGE.

Courteille, O., Bergin, R., Stockeld, D., Ponzer, S. & Fors, U. 2008. The use of a virtual patient case in an osce-based exam--a pilot study. *Medical Teacher,* 30**,** e66-76.

Croskerry, P. 2009a. Clinical cognition and diagnostic error: Applications of a dual process model of reasoning. *Advances in Health Sciences Education : Theory and Practice,* 14 Suppl 1**,** 27-35.

Croskerry, P. 2009b. A universal model of diagnostic reasoning. *Academic Medicine : Journal of the Association of American Medical Colleges,* 84**,** 1022-8.

Dacre, J. E. & Fox, R. A. 2000. How should we be teaching our undergraduates? *Annals of the Rheumatic Diseases,* 59**,** 662-7.

Dale, R. A., Sandoval, V. A., Hendricson, W. D. & Herbert, R. J. 1986. A comparison of simulation programs for endodontic review. *The Journal of biocommunication,* 13**,** 17-23.
De Jong, T. 2005. The guided discovery principle in multimedia learning. *In:* MAYER, R. E. (ed.) *The Cambridge handbook of multimedia learning.* Cambridge: University of Cambridge.
Decisionsim-Llc 2012. Decisionsim virtual patient authoring system. 2 ed.
Denzin, N. K. & Lincoln, Y. S. 2011. *The SAGE handbook of qualitative research,* Thousand Oaks, SAGE.
Devellis, R. F. 2003. *Scale development : Theory and applications,* Thousand Oaks, Calif. ; London, SAGE Publications, Inc.
Dixon, P. & Bortolussi, M. 2013. Construction, integration, and mind wandering in reading. *Canadian Journal of Experimental Psychology = Revue canadienne de psychologie experimentale,* 67**,** 1-10.
Dornan, T., Littlewood, S., Margolis, S. A., Scherpbier, A., Spencer, J. & Ypinazar, V. 2006. How can experience in clinical and community settings contribute to early medical education? A beme systematic review. *Medical Teacher,* 28**,** 3-18.
Dyrbye, L. N., Massie, F. S., Jr., Eacker, A., Harper, W., Power, D., Durning, S. J., Thomas, M. R., Moutier, C., Satele, D., Sloan, J. & Shanafelt, T. D. 2010. Relationship between burnout and professional conduct and attitudes among us medical students. *JAMA : the journal of the American Medical Association,* 304**,** 1173-80.
Edelbring, S. 2013. Research into the use of virtual patients is moving forward by zooming out. *Medical Education,* 47**,** 544-6.
Edelbring, S., Brostrom, O., Henriksson, P., Vassiliou, D., Spaak, J., Dahlgren, L. O., Fors, U. & Zary, N. 2012. Integrating virtual patients into courses: Follow-up seminars and perceived benefit. *Medical Education,* 46**,** 417-25.
Edelbring, S., Dastmalchi, M., Hult, H., Lundberg, I. E. & Dahlgren, L. O. 2011. Experiencing virtual patients in clinical learning: A phenomenological study. *Advances in Health Sciences Education : Theory and Practice,* 16**,** 331-45.
Ellaway, R. & Masters, K. 2008a. Amee guide 32: E-learning in medical education part 1: Learning, teaching and assessment. *Medical Teacher,* 30**,** 455-73.
Ellaway, R., Poulton, T., Fors, U., Mcgee, J. B. & Albright, S. 2008b. Building a virtual patient commons. *Medical Teacher,* 30**,** 170-4.
Ellaway, R. H. & Davies, D. 2011. Design for learning: Deconstructing virtual patient activities. *Medical Teacher,* 33**,** 303-10.
Elstein, A. S. 2009. Thinking about diagnostic thinking: A 30-year perspective. *Advances in Health Sciences Education : Theory and Practice,* 14 Suppl 1**,** 7-18.
Elstein, A. S., Shulman, L. S. & Sprafka, S. A. 1978. *Medical problem solving : An analysis of clinical reasoning,* Cambridge, Mass., Harvard University Press.
Ericsson KA. Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. *Academic Medicine : Journal of the Association of American Medical Colleges,* 2004;79.
Eurell, J. a. C., Lichtensteiger, C. A., Kingston, S. K., Diamond, N. A. & Miller, G. Y. 1999. Clinical cases as a teaching tool in veterinary histology. *J Vet Med Educ* 26**,** 1-6.

European Virtual Patients Project Development Team. 2010. *Virtual patient referatory* [Online]. Available: http://www.virtualpatients.eu/referatory [Accessed 12th August 2010].

Fern, E. F. 1982. The use of focus groups for idea generation: The effects of group size, acquaintanceship, and moderator on response quantity and quality. *Journal of Marketing Research,* 19**,** 1-13.

Finucane, P., Flannery, D., Mcgrath, D. & Saunders, J. 2013. Demographic attributes and knowledge acquisition among graduate-entry medical students. *Medical Teacher,* 35**,** 134-8.

Fischer, M. R., Schauer, S., Grasel, C., Baehring, T., Mandl, H., Gartner, R., Scherbaum, W. & Scriba, P. C. 1996. [casus model trial. A computer-assisted author system for problem-oriented learning in medicine]. *Zeitschrift fur arztliche Fortbildung,* 90**,** 385-9.

Flick, U. 2002. *An introduction to qualitative research,* London, SAGE.

Foo, J. L., Lobe, T. & Winer, E. 2009. A virtual reality environment for patient data visualization and endoscopic surgical planning. *Journal of laparoendoscopic & advanced surgical techniques. Part A,* 19 Suppl 1**,** S211-7.

Fors, U. G., Muntean, V., Botezatu, M. & Zary, N. 2009. Cross-cultural use and development of virtual patients. *Medical Teacher,* 31**,** 732-8.

Forsberg, E., Georg, C., Ziegert, K. & Fors, U. 2010. Virtual patients for assessment of clinical reasoning in nursing - a pilot study. *Nurse education today*. 757-62.

Fowler, J. 2008. Experiential learning and its facilitation. *Nurse education today,* 28**,** 427-33.

Franke, C., Holzum, A., Bohner, H., Baehring, T. & Ohmann, C. 2002. [computer-assisted case-based education in surgery]. *Der Chirurg; Zeitschrift fur alle Gebiete der operativen Medizen,* 73**,** 487-91.

Freeman, K. M., Thompson, S. F., Allely, E. B., Sobel, A. L., Stansfield, S. A. & Pugh, W. M. 2001. A virtual reality patient simulation system for teaching emergency response skills to u.S. Navy medical providers. *Prehospital and disaster medicine : the official journal of the National Association of EMS Physicians and the World Association for Emergency and Disaster Medicine in association with the Acute Care Foundation,* 16**,** 3-8.

Friedman, C. P. 1994. The research we should be doing. *Academic Medicine : Journal of the Association of American Medical Colleges,* 69**,** 455-7.

Friedman, C. P., France, C. L. & Drossman, D. D. 1991. A randomized comparison of alternative formats for clinical simulations. *Medical decision making : an international journal of the Society for Medical Decision Making,* 11**,** 265-72.

Gaufberg, E. H., Batalden, M., Sands, R. & Bell, S. K. 2010. The hidden curriculum: What can we learn from third-year medical student narrative reflections? *Academic Medicine : Journal of the Association of American Medical Colleges,* 85**,** 1709-16.

Gesundheit, N., Brutlag, P., Youngblood, P., Gunning, W. T., Zary, N. & Fors, U. 2009. The use of virtual patients to assess the clinical skills and reasoning of medical students: Initial insights on student acceptance. *Medical Teacher,* 31**,** 739-42.

Gibbs, G. 2007. *Analyzing qualitative data,* Los Angeles ; London, SAGE Publications.
Gill, C. J., Sabin, L. & Schmid, C. H. 2005. Why clinicians are natural bayesians. Bmj, 330, 1080-3.
Glaser, B. G. 1978. *Theoretical sensitivity : Advances in the methodology of grounded theory,* Mill Valley, Calif., Sociology Press.
Glaser, B. G. 1992. *Basics of grounded theory analysis : Emergence vs. Forcing,* Mill Valley, CA, Sociology Press.
Glaser, B. G. & Strauss, A. L. 1965. *Awareness of dying,* Chicago, Aldine Pub. Co.
Glaser, B. G. & Strauss, A. L. 1967. *Discovery of grounded theory : Strategies for qualitative research,* New York, Aldine Pub. Co.
Glaser, B. G. & Strauss, A. L. 1973. *The discovery of grounded theory : Strategies for qualitative research,* Chicago, Aldine Pub.
Goldenberg, M. J. 2006. On evidence and evidence-based medicine: Lessons from the philosophy of science. *Social science & medicine,* 62**,** 2621-32.
Gould, P. 2004. Creative commons ponders share options. *Nature,* 432**,** 137.
Green, M. L. & Ellis, P. J. 1997. Impact of an evidence-based medicine curriculum based on adult learning theory. *Journal of general internal medicine,* 12**,** 742-50.
Groves, M., Scott, I. & Alexander, H. 2002. Assessing clinical reasoning: A method to monitor its development in a pbl curriculum. *Medical Teacher,* 24**,** 507-15.
Hafferty, F. W. 1998. Beyond curriculum reform: Confronting medicine's hidden curriculum. *Academic Medicine : Journal of the Association of American Medical Colleges,* 73**,** 403-7.
Harless, W. G., Drennon, G. G., Marxer, J. J., Root, J. A. & Miller, G. E. 1971. Case: A computer-aided simulation of the clinical encounter. *Journal of medical education,* 46**,** 443-8.
Hawkes, N. 2012. Wellcome trust backs campaign for open access to research. *Bmj,* 344**,** e2653.
Heijne-Penninga, M., Kuks, J. B., Hofman, W. H. & Cohen-Schotanus, J. 2008. Influence of open- and closed-book tests on medical students' learning approaches. *Medical Education,* 42**,** 967-74.
Hemingway, A., Angell, C., Hartwell, H. & Heller, R. F. 2011. An emerging model for publishing and using open educational resources in public health. *Perspectives in public health,* 131**,** 38-43.
Hennink, M. M. 2010. International focus group research: A handbook for the health and social sciences. Cambridge: Cambridge University Press.
Herrle, S. R., Corbett, E. C., Jr., Fagan, M. J., Moore, C. G. & Elnicki, D. M. 2011. Bayes' theorem and the physical examination: Probability assessment and diagnostic decision making. *Academic Medicine : Journal of the Association of American Medical Colleges,* 86**,** 618-27.
Higgs, J., Jones, M. A. 2008. Clinical decision making and multiple problem spaces. . *In:* Higgs, J., Jones, M. A., Loftus, S., Christensen, N. (ed.) *Clinical reasoning in the health professions.* New York, NY: Butterworth Heinemann.
Holden, J. D. 2001. Hawthorne effects and research into professional practice. *Journal of evaluation in clinical practice,* 7**,** 65-70.

Huang, G., Reynolds, R. & Candler, C. 2007. Virtual patient simulation at us and canadian medical schools. *Academic Medicine : Journal of the Association of American Medical Colleges,* 82**,** 446-51.

Hubal, R. C., Kizakevich, P. N., Guinn, C. I., Merino, K. D. & West, S. L. 2000. The virtual standardized patient. Simulated patient-practitioner dialog for patient interview training. *Studies in health technology and informatics,* 70**,** 133-8.

Huhn, K., Black, L., Jensen, G. M. & Deutsch, J. E. 2011. Construct validity of the health science reasoning test. *Journal of allied health,* 40**,** 181-6.

Huwendiek, S. & De Leng, B. A. 2010. Virtual patient design and curricular integration evaluation toolkit. *Medical education,* 44**,** 519.

Huwendiek, S., De Leng, B. A., Zary, N., Fischer, M. R., Ruiz, J. G. & Ellaway, R. 2009a. Towards a typology of virtual patients. *Medical Teacher,* 31**,** 743-8.

Huwendiek, S., Reichert, F., Bosse, H.-M., De Leng, B. A., Van Der Vleuten, C. P. M., Haag, M., Hoffmann, G. F. & Tönshoff, B. 2009b. Design principles for virtual patients: A focus group study among students. *Medical Education,* 43**,** 580-588.

IBM. 2010. Spss statistics 19. IBM.

Illing, J. 2010. Thinking about research. *In:* Swanwick, T. (ed.) *Understanding medical education : Evidence, theory and practice.* Chichester: Wiley-Blackwell.

Institute for Innovative Technology in Medical Education (Iintime). 2012. *Computer-assisted learning in pediatrics program* [Online]. Available: http://www.med-u.org/virtual_patient_cases/clipp [Accessed 1st June 2012].

Issenberg, S. B., Mcgaghie, W. C., Petrusa, E. R., Lee Gordon, D. & Scalese, R. J. 2005. Features and uses of high-fidelity medical simulations that lead to effective learning: A beme systematic review. *Medical Teacher,* 27**,** 10-28.

Johnson, T. R., Lyons, R., Chuah, J. H., Kopper, R., Lok, B. C. & Cendan, J. C. 2013. Optimal leatning in a virtual patient simulation of cranial nerve palsies: The interaction between social learning context and student aptitude. *Medical Teacher,* 35**,** 876-84.

Jonassen, D. H., Lee, C. B., Yang, C. C. & Laffey, J. 2005. The collaboration principle. *In:* MAYER, R. E. (ed.) *The cambridge handbook of multimedia learning.* Cambridge: University of Cambridge.

Jupp, V. 2006. *The SAGE dictionary of social research methods,* London, SAGE.

Kang, M., Ragan, B. G. & Park, J. H. 2008. Issues in outcomes research: An overview of randomization techniques for clinical trials. *Journal of athletic training,* 43**,** 215-21.

Kassirer, J. P. 2010. Teaching clinical reasoning: Case-based and coached. *Academic Medicine : Journal of the Association of American Medical Colleges,* 85**,** 1118-24.

Kelle, U. 2007. The development of categories: Different approaches in grounded theory. *In:* Brant, A. & Charmaz, K. (eds.) *The SAGE handbook of grounded theory.* London: SAGE.

Kelly, E. & Nisker, J. 2010. Medical students' first clinical experiences of death. *Medical Education,* 44**,** 421-8.

Kelly, W. F., Papp, K. K., Torre, D. & Hemmer, P. A. 2012. How and why internal medicine clerkship directors use locally developed, faculty-written

examinations: Results of a national survey. *Academic Medicine : Journal of the Association of American Medical Colleges,* 87**,** 924-30.
Kennedy, T. J. & Lingard, L. A. 2006. Making sense of grounded theory in medical education. *Medical Education,* 40**,** 101-8.
Khan, N., Coppola, W., Rayne, T. & Epstein, O. 2009. Medical student access to multimedia devices: Most have it, some don't and what's next? *Informatics for health & social care,* 34**,** 100-5.
Killingsworth, M. A. & Gilbert, D. T. 2010. A wandering mind is an unhappy mind. *Science,* 330**,** 932-932.
Kim, S., Phillips, W. R., Pinsky, L., Brock, D., Phillips, K. & Keary, J. 2006. A conceptual framework for developing teaching cases: A review and synthesis of the literature across disciplines. *Medical Education,* 40**,** 867-76.
Kinzie, M. B., Cohn, W. F., Julian, M. F. & Knaus, W. A. 2002. A user-centered model for web site design: Needs assessment, user interface design, and rapid prototyping. *Journal of the American Medical Informatics Association : JAMIA,* 9**,** 320-30.
Kirkpatrick, D. L. 1967. Evaluation of training. *In:* Craig, R. & Mittel, I. (eds.) *Training and development handbook* New York: McGraw Hill.
Kiss, B., Benedek, B., Szijarto, G., Csukly, G., Simon, L. & Takacs, B. 2004. Virtual patient: A photo-real virtual human for vr-based therapy. *Studies in health technology and informatics,* 98**,** 154-6.
Kitzinger, J. 1995. Qualitative research. Introducing focus groups. *Bmj,* 311**,** 299-302.
Kneebone, R. L., Scott, W., Darzi, A. & Horrocks, M. 2004. Simulation and clinical practice: Strengthening the relationship. *Medical Education,* 38**,** 1095-102.
Kolb, D. A. 1984. *Experiential learning : Experience as the source of learning and development,* Englewood Cliffs, N.J., Prentice-Hall.
Krueger, R. A. & Casey, M. A. 2000. *Focus groups : A practical guide for applied research,* Thousand Oaks, Calif. ; London, SAGE Publications.
Krueger, R. A. & Casey, M. A. 2009. *Focus groups : A practical guide for applied research,* Los Angeles, SAGE.
Kuper, A., Reeves, S. & Levinson, W. 2008a. An introduction to reading and appraising qualitative research. *Bmj,* 337**,** a288-a288.
Kuper, A., Reeves, S. & Levinson, W. 2008b. An introduction to reading and appraising qualitative research. *Bmj,* 337**,** a288.
Lechner, S. K., Thomas, G. A., Bradshaw, M. & Lechner, K. M. 2001. Planning oral rehabilitation: Case-based computer assisted learning in clinical dentistry. *British dental journal,* 191**,** 152-6.
Lee, J. E. & Ryan-Wenger, N. 1997. The "think aloud" seminar for teaching clinical reasoning: A case study of a child with pharyngitis. *Journal of pediatric health care : official publication of National Association of Pediatric Nurse Associates & Practitioners,* 11**,** 101-10.
Lee, M. Y., Albright, S., O'leary, L., Terkla, D. G. & Wilson, N. 2008. Expanding the reach of health sciences education and empowering others: The opencourseware initiative at tufts university. *Medical Teacher,* 30**,** 159-63.
Lempert, L. B. 2007. Asking questions of the data: Memo writing in the grounded theory tradition. *In:* Bryant, A. & Charmaz, K. (eds.) *The SAGE handbook of grounded theory.* London: SAGE.

Lewis, J. & Ritchie, J. 2003. *Qualitative research practice : A guide for social science students and researchers,* London ; Thousand Oaks, Calif., SAGE Publications.
Lin, C. C., Wu, W. C., Liaw, H. T. & Liu, C. C. 2012. Effectiveness of a virtual patient program in a psychiatry clerkship. *Medical Education,* 46**,** 1111-2.
Lingard, L., Albert, M. & Levinson, W. 2008. Grounded theory, mixed methods, and action research. *Bmj,* 337**,** a567.
Lingard, L. & Kennedy, T. J. 2010. Qualitative research methods in medical education. *In:* Swanwick, T. (ed.) *Understanding medical education : Evidence, theory and practice.* Chichester: Wiley-Blackwell.
Ludwig, D. S. & Kabat-Zinn, J. 2008. Mindfulness in medicine. *JAMA : the journal of the American Medical Association,* 300**,** 1350-2.
Lyon, H. C., Jr., Healy, J. C., Bell, J. R., O'donnell, J. F., Shultz, E. K., Wigton, R. S., Hirai, F. & Beck, J. R. 1991. Findings from an evaluation of planalyzer's double cross-over trials of computer-based, self-paced, case-based programs in anemia and chest pain diagnosis. *Proceedings / the Annual Symposium on Computer Application in Medical Care. Symposium on Computer Applications in Medical Care***,** 88-93.
Malterud, K. 2001. Qualitative research: Standards, challenges, and guidelines. *The Lancet,* 358**,** 483-488.
Mayer, R. E. 2005. *The cambridge handbook of multimedia learning,* Cambridge, University of Cambridge.
Mayer, R. E. 2010. Applying the science of learning to medical education. *Medical Education,* 44**,** 543-9.
Mayor, S. 2013. Research council failed to clearly communicate its open access policy, say peers. *Bmj,* 346.
Mcgee, J. B., Neill, J., Goldman, L. & Casey, E. 1998. Using multimedia virtual patients to enhance the clinical curriculum for medical students. *Studies in health technology and informatics,* 52 Pt 2**,** 732-5.
MedBiquitous. 2011. *Virtual patient implementers* [Online]. Available: http://www.medbiq.org/virtual_patient/implementers [Accessed 21st November 2012].
MedBiquitous Virtual Patient Working Group. 2010. *MedBiquitous virtual patient player specifications and description document v.1.0* [Online]. Medibiquitous. Available: http://www.medbiq.org/working_groups/virtual_patient/VirtualPatientPlayerSpecification.pdf [Accessed 20th April 2011].
Moher, D., Schulz, K. F. & Altman, D. G. 2001. The consort statement: Revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet,* 357**,** 1191-4.
Montgomery, A. A., Peters, T. J. & Little, P. 2003. Design, analysis and presentation of factorial randomised controlled trials. *BMC medical research methodology,* 3**,** 26.
Morrison, J. M., Sullivan, F., Murray, E. & Jolly, B. 1999. Evidence-based education: Development of an instrument to critically appraise reports of educational interventions. *Medical Education,* 33**,** 890-3.
Morse, J. M. 2009. Tussles, tensions, and resolutions. *In:* Morse, J. M., Stern, P. N., Corbin, J. M., Bowers, B., Charmaz, K. & Clarke, A. E. (eds.) *Developing*

grounded theory. The second generation. Walnut Creek, CA.: Left Coast Press.
Mullaney, T. P., Smith, T. A., Duell, R. C. & Kaplan, A. 1976. Four-phase study of computer-assisted and slide-tape methods of stimulating clinical endodontic problems. *Journal of dental education,* 40**,** 681-7.
Newble, D., Norman, G., Van Der Vleuten, C.P. 2000. Assessing clinical reasoning. *In:* JONES, J. H. M. (ed.) *Clinical reasoning in the health professions.* 2nd ed. Oxford: Butterworth-Heinemann.
Newble, D. I. & Jaeger, K. 1983. The effect of assessments and examinations on the learning of medical students. *Medical Education,* 17**,** 165-71.
Norman, G. 2005. Research in clinical reasoning: Past history and current trends. *Medical Education,* 39**,** 418-27.
Norman, G. & Eva, K. W. 2008. *Quantitative research methods in medical education,* Edinburgh, ASME.
Northern Ontario School of Medicine. 2010. *Pine project* [Online]. Available: http://www.pine.nosm.ca/pine/ [Accessed 2nd May 2010].
O'brien, M. 1993. Social research and sociology. *In:* Gilbert, G. N. (ed.) *Researching social life.* London: SAGE.
Orr, K. K. 2007. Integrating virtual patients into a self-care course. *American journal of pharmaceutical education,* 71**,** 30.
Orton, E. & Mulhausen, P. 2008. E-learning virtual patients for geratric education. *Gerontology & geriatrics education,* 28**,** 73-88.
Page, G., Bordage, G. & Allen, T. 1995. Developing key-feature problems and examinations to assess clinical decision-making skills. *Academic Medicine : Journal of the Association of American Medical Colleges,* 70**,** 194-201.
Patton, M. Q. 2002. *Qualitative research and evaluation methods,* Thousand Oaks, Calif., SAGE Publications.
Pettersson, S., Moller, S., Svenungsson, E., Gunnarsson, I. & Welin Henriksson, E. 2010. Women's experience of sle-related fatigue: A focus group interview study. *Rheumatology,* 49**,** 1935-42.
Pope, C., Ziebland, S. & Mays, N. 2007. Analysing qualitative data. *In:* POPE, C., MAYS, N. & POPAY, J. (eds.) *Synthesizing qualitative and quantitative health evidence : A guide to methods.* Maidenhead, England ; New York: Open University Press, McGraw Hill Education.
Posel, N., Fleiszer, D. & Shore, B. M. 2009. 12 tips: Guidelines for authoring virtual patient cases. *Medical Teacher,* 31**,** 701-8.
Qsr-International-Ltd. 2011. Nvivo qualitative data analysis software. Version 9.0.
Round, A. P. 1999. Teaching clinical reasoning--a preliminary controlled study. *Medical Education,* 33**,** 480-3.
Round, J., Conradi, E. & Poulton, T. 2009. Improving assessment with virtual patients. *Medical Teacher,* 31**,** 759-63.
Ruderich, F., Bauch, M., Haag, M., Heid, J., Leven, F. J., Singer, R., Geiss, H. K., Junger, J. & Tonshoff, B. 2004. Campus--a flexible, interactive system for web-based, problem-based learning in health care. *Studies in health technology and informatics,* 107**,** 921-5.
Salas, E., Wilson, K. A., Burke, C. S. & Priest, H. A. 2005. Using simulation-based training to improve patient safety: What does it take? *Joint Commission*

journal on quality and patient safety / Joint Commission Resources, 31, 363-71.
Sandoval, V. A., Dale, R. A., Hendricson, W. D. & Alexander, J. B. 1987. A comparison of four simulation and instructional methods for endodontic review. *Journal of dental education,* 51, 532-8.
Sanyer, O. N. & Fortenberry, K. 2013. Using mindfulness techniques to improve difficult clinical encounters. *American family physician,* 87, 402.
Schardt, C., Adams, M. B., Owens, T., Keitz, S. & Fontelo, P. 2007. Utilization of the pico framework to improve searching pubmed for clinical questions. *BMC medical informatics and decision making,* 7, 16.
Schmidt, H. G., Norman, G. R. & Boshuizen, H. P. 1990. A cognitive perspective on medical expertise: Theory and implication. *Academic Medicine : Journal of the Association of American Medical Colleges,* 65, 611-21.
Schon, D. A. 1987. *Educating the reflective practitioner: Towards a new design for teaching and learning in the professions,* San Fransisco, CA, Jossey-Bass.
Schuwirth, L. 2009. Is assessment of clinical reasoning still the holy grail? *Medical Education,* 43, 298-300.
Scott, S. 2013. The researcher of the future...Makes the most of social media. *Lancet,* 381 Suppl 1, S5-6.
Sedlmeier, P. & Gigerenzer, G. 2001. Teaching bayesian reasoning in less than two hours. *Journal of experimental psychology. General,* 130, 380-400.
Shachak, A., Ophir, R. & Rubin, E. 2005. Applying instructional design theories to bioinformatics education in microarray analysis and primer design workshops. *Cell biology education,* 4, 199-206.
Shacklock, G. & Smyth, J. 1998. *Being reflexive in critical educational and social research,* London, Falmer Press.
Shapiro Institute for Education and Research & Beth Israel Deaconess Medical Centre. 2004. *Virtual patient open courseware* [Online]. Harvard: Harvard Medical School. Available: http://www.mycourses.med.harvard.edu/ocw_rbview.asp?course=VP [Accessed 22nd December 2010].
Sieper, J. & Rudwaleit, M. 2005. Early referral recommendations for ankylosing spondylitis (including pre-radiographic and radiographic forms) in primary care. *Annals of the rheumatic diseases,* 64, 659-63.
Smallwood, J., Mrazek, M. D. & Schooler, J. W. 2011. Medicine for the wandering mind: Mind wandering in medical practice. *Medical Education,* 45, 1072–1080.
Smothers, V., Ellaway, R. & Balasubramaniam, C. 2008. Evip: Sharing virtual patients across europe. *AMIA Annu Symp Proc*, 1140.
Sobral, D. T. 1995. Diagnostic ability of medical students in relation to their learning characteristics and preclinical background. *Medical Education,* 29, 278-82.
St Georges Univeristy E-Learning Unit. 2010. *Virtual patient example pages* [Online]. Available: http://www.elu.sgul.ac.uk/virtualpatients/examples.htm [Accessed 1st May 2010].
Stern, P. N. 1995. Eroding grounded theory. *In:* J.M., M. (ed.) *Critical issues in qulaitative research methods.* Thousand Oaks, CA: SAGE.
Stewart, D. & Shamdasani, R. 1990. *focus groups: Theory and practice. Applied social research methods series, volume 20.,* CA, SAGE Publications.

Strauss, A. L. 1987. *Qualitative analysis for social scientists,* Cambridge, NY, Cambridge University Press.

Strauss, A. L. & Corbin, J. M. 1990a. *Basics of qualitative research : Grounded theory procedures and techniques,* Newbury Park, Calif., SAGE Publications.

Strauss, A. L. & Corbin, J. M. 1990b. Basics of qualitative research : Grounded theory procedures and techniques.

Strauss, A. L. & Corbin, J. M. 1998a. *Basics of qualitative research : Techniques and procedures for developing grounded theory,* Thousand Oaks, Calif., SAGE Publications.

Strauss, A. L. & Corbin, J. M. 1998b. Basics of qualitative research : Techniques and procedures for developing grounded theory.

Sweller, J. 2005. Implications of cognitive load theory for multimedia learning. *In:* MAYER, R. E. (ed.) *The cambridge handbook of multimedia learning.* Cambridge: University of Cambridge.

Tabachnick, B. G. & Fidell, L. S. 2007. *Using multivariate statistics,* Boston ; London, Pearson/A & B.

Tabachnick, B. G. & Fidell, L. S. 2012. *Using multivariate statistics,* Boston, Mass. ; London, Pearson.

Taitz, J., Wyeth, B., Lennon, R., Torre, P. D., Yen, T., Harrison, B. & Cattell, M. 2006. Effect of the introduction of a lumbar-puncture sticker and teaching manikin on junior staff documentation and performance of paediatric lumbar punctures. *Quality & safety in health care,* 15**,** 325-8.

Taylor, D., Patel, V., Cohen, D., Aggarwal, R., Kerr, K., Sevdalis, N., Batrick, N. & Darzi, A. 2011. Single and multi-user virtual patient design in the virtual world. *Studies in health technology and informatics,* 163**,** 650-2.

Teunissen, P. W., Boor, K., Scherpbier, A. J., Van Der Vleuten, C. P., Van Diemen-Steenvoorde, J. A., Van Luijk, S. J. & Scheele, F. 2007a. Attending doctors' perspectives on how residents learn. *Medical Education,* 41**,** 1050-8.

Teunissen, P. W., Scheele, F., Scherpbier, A. J., Van Der Vleuten, C. P., Boor, K., Van Luijk, S. J. & Van Diemen-Steenvoorde, J. A. 2007b. How residents learn: Qualitative evidence for the pivotal role of clinical activities. *Medical Education,* 41**,** 763-70.

Triola, M., Feldman, H., Kalet, A. L., Zabar, S., Kachur, E. K., Gillespie, C., Anderson, M., Griesser, C. & Lipkin, M. 2006. A randomized trial of teaching clinical skills using virtual and live standardized patients. *Journal of general internal medicine,* 21**,** 424-9.

Triola, M. M., Campion, N., Mcgee, J. B., Albright, S., Greene, P., Smothers, V. & Ellaway, R. 2007. An xml standard for virtual patients: Exchanging case-based simulations in medical education. *AMIA ... Annual Symposium proceedings / AMIA Symposium. AMIA Symposium***,** 741-5.

Tworek, J., Coderre, S., Wright, B. & Mclaughlin, K. 2010. Virtual patients: Ed-2 band-aid or valuable asset in the learning portfolio? *Academic Medicine : Journal of the Association of American Medical Colleges,* 85**,** 155-8.

Unesco 2011. *Guidelines for open educational resources (oer) in higher education*, UNESCO.

Us National Library of Medicine & National Institute of Health. 2012. *Open access subset* [Online]. Available:

http://www.ncbi.nlm.nih.gov/pmc/tools/openftlist/ [Accessed 11th November 2012].

Voelker, R. 2003. Virtual patients help medical students link basic science with clinical care. *JAMA : the journal of the American Medical Association,* 290**,** 1700-1.

Watling, C. J. & Lingard, L. 2012. Grounded theory in medical education research: Amee guide no. 70. *Medical Teacher,* 34**,** 850-61.

Wilson, A. S., Goodall, J. E., Ambrosini, G., Carruthers, D. M., Chan, H., Ong, S. G., Gordon, C. & Young, S. P. 2006. Development of an interactive learning tool for teaching rheumatology--a simulated clinical case studies program. *Rheumatology,* 45**,** 1158-61.

Windish, D. M., Price, E. G., Clever, S. L., Magaziner, J. L. & Thomas, P. A. 2005. Teaching medical students the important connection between communication and clinical reasoning. *Journal of general internal medicine,* 20**,** 1108-13.

Wolpaw, T., Papp, K. K. & Bordage, G. 2009. Using snapps to facilitate the expression of clinical reasoning and uncertainties: A randomized comparison group trial. *Academic Medicine : Journal of the Association of American Medical Colleges,* 84**,** 517-24.

World Wide Web Consortium. 2008. *Extensible markup language (xml) 1.0* [Online]. W3C. Available: http://www.w3.org/TR/REC-xml/ - sec-origin-goals [Accessed 6th April 2011].

Yang, R. L., Hashimoto, D. A., Predina, J. D., Bowens, N. M., Sonnenberg, E. M., Cleveland, E. C., Lawson, C., Morris, J. B. & Kelz, R. R. 2013. The virtual-patient pilot: Testing a new tool for undergraduate surgical education and assessment. *Journal of surgical education,* 70**,** 394-401.

Yanko, M. R. & Spalek, T. M. 2012. Mind-wandering impairs driving performance. *Canadian Journal of Experimental Psychology-Revue Canadienne De Psychologie Experimentale,* 66**,** 296-296.

Yardley, S. & Dornan, T. 2012. Kirkpatrick's levels and education 'evidence'. *Medical Education,* 46**,** 97-106.

Zary, N., Johnson, G., Boberg, J. & Fors, U. G. 2006. Development, implementation and pilot evaluation of a web-based virtual patient case simulation environment--web-sp. *BMC medical education,* 6**,** 10.

Zary, N., Johnson, G. & Fors, U. 2009. Web-based virtual patients in dentistry: Factors influencing the use of cases in the web-sp system. *European journal of dental education : official journal of the Association for Dental Education in Europe,* 13**,** 2-9.

Zillich, A. J., Ackermann, R. T., Stump, T. E., Ambuehl, R. J., Downs, S. M., Holmes, A. M., Katz, B. & Inui, T. S. 2008. An evaluation of educational outreach to improve evidence-based prescribing in medicaid: A cautionary tale. *Journal of evaluation in clinical practice,* 14**,** 854-60.

Section 9. Appendices

Appendix 1. Ethics approval

This research has received written ethics approval from the Warwick Medical School Biomedical Research Ethics committee (3.8.10), Keele School of Medicine Ethics Committee, and the NHS National Research Ethics Service.

Documentation is shown below and on the following page.

03 August 2010

Dr James Bateman
40 Grange Road
Halesowen
B63 3EQ

Dear James

The effect of virtual patient design and feedback on the achievement of learning outcomes in undergraduate musculoskeletal medicine: a prospective mixed methods study of virtual patients

Thank you for submitting revised documentation in support of the above study submitted for review to the University of Warwick's Biomedical Research Ethics Sub-Committee (BREC) for Chair's Approval.

I am pleased to confirm that your revisions address the Committee's concerns. I therefore confirm your study has been awarded full approval and may commence.

May I take this opportunity to wish all success with the project and to remind you that any substantial amendments require prior approval from the BREC Chair. It would also be appreciated if the Committee could receive an end of study report by email to Krysia Saul.

Yours sincerely,

Jane Barlow
Chair
Biomedical Research
Ethics Sub-Committee

Copy:
David Davies, Academic Supervisor

Biomedical Research Ethics Subcommittee
Enquiries: Krysia Saul
Tel: 02476-573163
Email: krysia.saul@warwick.ac.uk

NHS
National Research Ethics Service

West Midlands Research Ethics Committee
Osprey House
Albert Street
Redditch
Worcestershire, B97 4DE
anne.mccullough@westmidlands.nhs.uk
Chairman: Mr Paul Hamilton

Telephone: 01527 587573
Facsimile: 01527 587501

14 August 2009

Mr James Bateman

Dear Mr Bateman

Arc Educational Research Fellowship Project

Thank you for seeking the Committee's advice about the above project. You provided the following documents for consideration:

Research Ethics Proposal Summary

This document was considered by the West Midlands Research Ethics Committee Chairman.

I enclose a copy of our leaflet, "Defining Research", which explains how we differentiate research from other activities. The Chairman has advised that the project is not considered to be research according to this guidance. Therefore it does not require ethical review by a NHS Research Ethics Committee.

You should check with the NHS care organisation (if applicable) what other review arrangements or sources of advice apply to projects of this type. Guidance may be available from the clinical governance office.

You may wish to check whether the project should be reviewed by the ethics committee within your own institution.

This letter should not be interpreted as giving a form of ethical approval or any endorsement of the project, but it may be provided to a journal or other body as evidence that ethical approval is not required under NHS research governance arrangements.

However, if you, your sponsor/funder or any NHS organisation feels that the project should be managed as research and/or that ethical review by a NHS REC is essential, please write setting out your reasons and we will be pleased to consider further.

Where NHS organisations have clarified that a project is not to be managed as research, the Research Governance Framework states that it should not be presented as research within the NHS.

Yours sincerely

Maxine North

Mrs Anne McCullough
Committee Co-ordinator

E-mail: anne.mccullough@westmidlands.nhs.uk

Enclosure: *NRES leaflet – "Defining Research"*

Appendix 2 Literature review: papers selected for full text review.

1: Botezatu M, Hult H, Tessma MK, Fors U. Virtual patient simulation: knowledge gain or knowledge loss? Med Teach. 2010;32(7):562-8.

2: Huwendiek S, de Leng BA. Virtual patient design and curricular integration evaluation toolkit. Med Educ. 2010 May;44(5):519.

3: Tan ZS, Mulhausen PL, Smith SR, Ruiz JG. Virtual patients in geriatric education. Gerontol Geriatr Educ. 2010 Apr;31(2):163-73.

4: Lehmann R, Bosse HM, Huwendiek S. Blended learning using virtual patients and skills laboratory training. Med Educ. 2010 May;44(5):521-2.

5: Ninomiya S, Tokaji M, Tokumine A, Kurosaki T. Virtual patient simulator for the perfusion resource management drill. J Extra Corpor Technol. 2009 Dec;41(4):206-12.

6: Tworek J, Coderre S, Wright B, McLaughlin K. Virtual patients: ED-2 band-aid or valuable asset in the learning portfolio? Acad Med. 2010 Jan;85(1):155-8.

7: Horstmann M, Renninger M, Hennenlotter J, Horstmann CC, Stenzl A. Blended E-learning in a Web-based virtual hospital: a useful tool for undergraduate education in urology. Educ Health (Abingdon). 2009 Aug;22(2):269. Epub 2009 Jul 30.

8: Graham JM, Brinson M, Magtibay LV, Regan B, Lazar EJ. Virtual patient safety rounds: one hospital system's approach to sharing knowledge. J Healthc Qual. 2009 Sep-Oct;31(5):48-52.

9: Round J, Conradi E, Poulton T. Improving assessment with virtual patients. Med Teach. 2009 Aug;31(8):759-63.

10: Round J, Conradi E, Poulton T. Training staff to create simple interactive virtual patients: the impact on a medical and healthcare institution. Med Teach. 2009 Aug;31(8):764-9.

11: Poulton T, Conradi E, Kavia S, Round J, Hilton S. The replacement of 'paper' cases by interactive online virtual patients in problem-based learning. Med Teach. 2009 Aug;31(8):752-8.

12: Huwendiek S, De leng BA, Zary N, Fischer MR, Ruiz JG, Ellaway R. Towards a typology of virtual patients. Med Teach. 2009 Aug;31(8):743-8.

13: Gesundheit N, Brutlag P, Youngblood P, Gunning WT, Zary N, Fors U. The use of virtual patients to assess the clinical skills and reasoning of medical students: initial insights on student acceptance. Med Teach. 2009 Aug;31(8):739-42. PubMed PMID: 19811211.

14: Fors UG, Muntean V, Botezatu M, Zary N. Cross-cultural use and development of virtual patients. Med Teach. 2009 Aug;31(8):732-8.

15: Dewhurst D, Borgstein E, Grant ME, Begg M. Online virtual patients - A driver for change in medical and healthcare professional education in developing countries? Med Teach. 2009 Aug;31(8):721-4.

16: Conradi E, Kavia S, Burden D, Rice A, Woodham L, Beaumont C, Savin-Baden M, Poulton T. Virtual patients in a virtual world: Training paramedic students for practice. Med Teach. 2009 Aug;31(8):713-20.

17: Campbell G, Miller A, Balasubramaniam C. The role of intellectual property in creating, sharing and repurposing virtual patients. Med Teach. 2009 Aug;31(8):709-12.

18: Ellaway RH, Poulton T, Smothers V, Greene P. Virtual patients come of age. Med Teach. 2009 Aug;31(8):683-4.

19: Kim KJ, Han J, Park IeB, Kee C. Medical education in Korea: the e-learning consortium. Med Teach. 2009 Sep;31(9):e397-401.

20: Zary N, Hege I, Heid J, Woodham L, Donkers J, Kononowicz AA. Enabling interoperability, accessibility and reusability of virtual patients across Europe - design and implementation. Stud Health Technol Inform. 2009;150:826-30.

21: Kononowicz AA, Heid J, Donkers J, Hege I, Woodham L, Zary N. Development and validation of strategies to test for interoperability of virtual patients. Stud Health Technol Inform. 2009;150:185-9.

22: Bardella IJ. Are virtual patients a valid assessment tool? Fam Med. 2009 Sep;41(8):594.

23: Kenny PG, Parsons TD, Rothbaum B, Difede J, Reger G, Rizzo A. Optimizing Clinical Training for the Treatment of PTSD Using Virtual Patients. Stud Health Technol Inform. 2009;144:264-8.

24: Berman N, Fall LH, Smith S, Levine DA, Maloney CG, Potts M, Siegel B,Foster-Johnson L. Integration strategies for using virtual patients in clinical clerkships. Acad Med. 2009 Jul;84(7):942-9.

25: Posel N, Fleiszer D, Shore BM. 12 Tips: Guidelines for authoring virtual patient cases. Med Teach. 2009 Aug;31(8):701-8.

26: Huwendiek S, Reichert F, Bosse HM, de Leng BA, van der Vleuten CP, Haag M, Hoffmann GF, Tönshoff B. Design principles for virtual patients: a focus group study among students. Med Educ. 2009 Jun;43(6):580-8.

27: Nirenburg S, McShane M, Beale S, Jarrell B, Fantry G. Integrating cognitive simulation into the Maryland virtual patient. Stud Health Technol Inform. 2009;142:224-9.

28: Cook DA, Triola MM. Virtual patients: a critical literature review and proposed next steps. Med Educ. 2009 Apr;43(4):303-11. Review.

29: Zary N, Johnson G, Fors U. Web-based virtual patients in dentistry: factors influencing the use of cases in the Web-SP system. Eur J Dent Educ. 2009 Feb;13(1):2-9.

30: McConnell H, Pardy A. Virtual patient simulation for prevention of medical error: beyond just technical upskilling. World Hosp Health Serv. 2008;44(3):36-9. PubMed PMID: 19181024.

31: Deladisma AM, Gupta M, Kotranza A, Bittner JG 4th, Imam T, Swinson D, Gucwa A, Nesbit R, Lok B, Pugh C, Lind DS. A pilot study to integrate an immersive virtual patient with a breast complaint and breast examination simulator into a surgery clerkship. Am J Surg. 2009 Jan;197(1):102-6.

32: Smothers V, Ellaway R, Balasubramaniam C. EViP: sharing virtual patients across Europe. AMIA Annu Symp Proc. 2008 Nov 6:1140.

33: Sumner W 2nd, Hagen MD. Value of information in virtual patient performance evaluations. AMIA Annu Symp Proc. 2008 Nov 6:1149.

34: Sumner W 2nd, Hagen MD. Value of information in virtual patients portraying pharyngitis. AMIA Annu Symp Proc. 2008 Nov 6:1150.

35: Critchley LA, Wong JW, Leung JY. Virtual patients and undergraduate anaesthesia teaching. Med Educ. 2008 Oct 1.

36: Hooper LM, Weinfurt KP, Cooper LA, Mensh J, Harless W, Kuhajda MC, Epstein SA. Virtual standardized patients: an interactive method to examine variation in depression care among primary care physicians. Prim Health Care Res Dev. 2008 Oct 1;9(4):257-268.

37: Triola MM, Campion N, McGee JB, Albright S, Greene P, Smothers V, Ellaway R. An XML standard for virtual patients: exchanging case-based simulations in medical education. AMIA Annu Symp Proc. 2007 Oct 11:741-5.

38: Dev P, Heinrichs WL, Youngblood P, Kung S, Cheng R, Kusumoto L, Hendrick A. Virtual patient model for multi-person virtual medical environments. AMIA Annu Symp Proc. 2007 Oct 11:181-5.

39: Vukanovic-Criley JM, Boker JR, Criley SR, Rajagopalan S, Criley JM. Using virtual patients to improve cardiac examination competency in medical students. Clin Cardiol. 2008 Jul;31(7):334-9.

40: Courteille O, Bergin R, Stockeld D, Ponzer S, Fors U. The use of a virtual patient case in an OSCE-based exam--a pilot study. Med Teach. 2008;30(3):e66-76.

41: Ellaway R, Poulton T, Fors U, McGee JB, Albright S. Building a virtual patient commons. Med Teach. 2008;30(2):170-4.

42: Deladisma AM, Johnsen K, Raij A, Rossen B, Kotranza A, Kalapurakal M, Szlam S, Bittner JG 4th, Swinson D, Lok B, Lind DS. Medical student satisfaction using a virtual patient system to learn history-taking communication skills. Stud Health Technol Inform. 2008;132:101-5.

43: Waldmann UM, Gulich MS, Zeitler HP. Virtual patients for assessing medical students--important aspects when considering the introduction of a new assessment format. Med Teach. 2008 Feb;30(1):17-24. PubMed PMID: 18278646.

44: Orton E, Mulhausen P. E-learning virtual patients for geratric education. Gerontol Geriatr Educ. 2008;28(3):73-88.

45: Ghosh S, Young DL, Gadkar KG, Wennerberg L, Basu K. Towards optimal virtual patients: an online adaptive control approach. Conf Proc IEEE Eng Med Biol Soc.2007;2007:3292-5.

46: Orr KK. Integrating virtual patients into a self-care course. Am J Pharm Educ. 2007 Apr 15;71(2):30. Pu

47: Deladisma AM, Cohen M, Stevens A, Wagner P, Lok B, Bernard T, Oxendine C, Schumacher L, Johnsen K, Dickerson R, Raij A, Wells R, Duerson M, Harper JG, Lind DS; Association for Surgical Education. Do medical students respond empathetically to a virtual patient? Am J Surg. 2007 Jun;193(6):756-60.

48: Huang G, Reynolds R, Candler C. Virtual patient simulation at US and Canadian medical schools. Acad Med. 2007 May;82(5):446-51.

49: Vash JH, Yunesian M, Shariati M, Keshvari A, Harirchi I. Virtual patients in undergraduate surgery education: a randomized controlled study. ANZ J Surg. 2007 Jan-Feb;77(1-2):54-9. PubMed PMID: 17295822.

50: Triola M, Feldman H, Kalet AL, Zabar S, Kachur EK, Gillespie C, Anderson M, Griesser C, Lipkin M. A randomized trial of teaching clinical skills using virtual and live standardized patients. J Gen Intern Med. 2006 May;21(5):424-9.

51: Zary N, Johnson G, Boberg J, Fors UG. Development, implementation and pilot evaluation of a Web-based Virtual Patient Case Simulation environment--Web-SP. BMC Med Educ. 2006 Feb 21;6:10.

52: Dickerson R, Johnsen K, Raij A, Lok B, Stevens A, Bernard T, Lind DS. Virtual patients: assessment of synthesized versus recorded speech. Stud Health Technol Inform. 2006;119:114-9.

53: Walsh K. Virtual patients get real. Med Educ. 2005 Nov;39(11):1153-4.

54: Simo A, Cavazza M, Kijima R. Virtual patients in clinical medicine. Stud Health Technol Inform. 2004;98:353-9.

55: Ruderich F, Bauch M, Haag M, Heid J, Leven FJ, Singer R, Geiss HK, Jünger J, Tönshoff B. CAMPUS--a flexible, interactive system for web-based, problem-based learning in health care. Stud Health Technol Inform. 2004;107(Pt 2):921-5.

56: Marco AP. The virtual patient encounter--units of service in the electronic age. Physician Exec. 2004 May-Jun;30(3):32-4.

57: Voelker R. Virtual patients help medical students link basic science with clinical care. JAMA. 2003 Oct 1;290(13):1700-1.

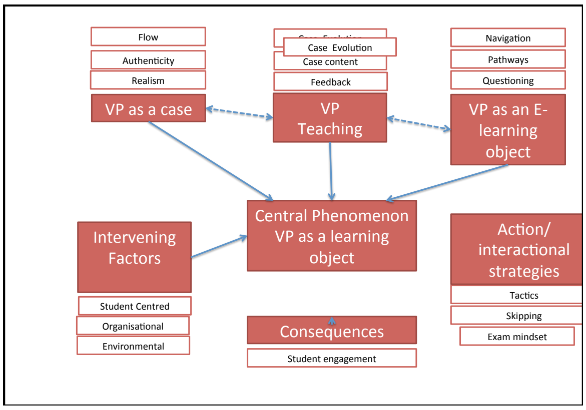
58: Malamateniou F, Vassilacopoulos G. Developing a virtual patient record using XML and web-based workflow technologies. Int J Med Inform. 2003 Jul;70(2-3):131-9.

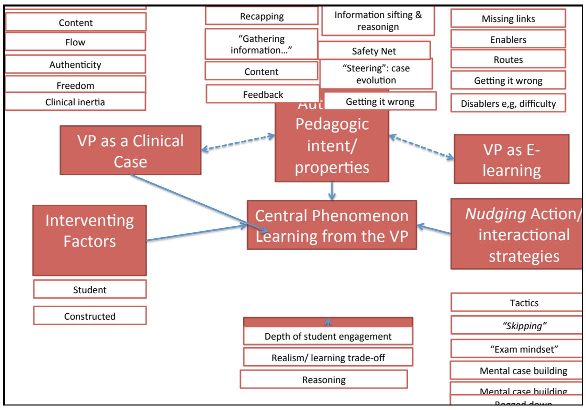
59: Bearman M. Is virtual the same as real? Medical students' experiences of a virtual patient. Acad Med. 2003 May;78(5):538-45.

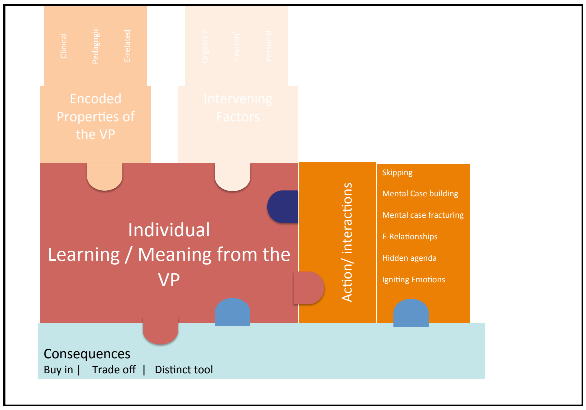
60: Bearman M, Cesnik B. Comparing student attitudes to different models of the same virtual patient. Stud Health Technol Inform. 2001;84(Pt 2):1004-8.

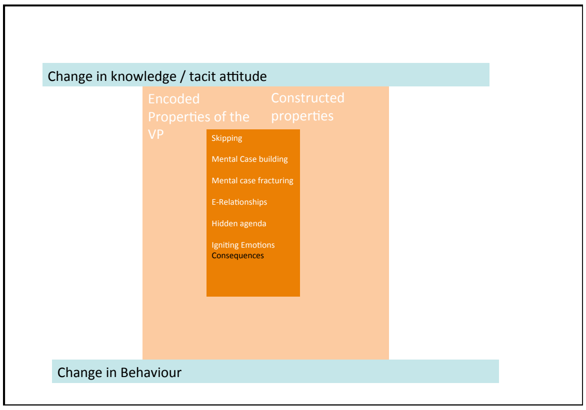
61: Bearman M, Cesnik B, Liddell M. Random comparison of 'virtual patient' models in the context of teaching clinical communication skills. Med Educ. 2001 Sep;35(9):824-32.

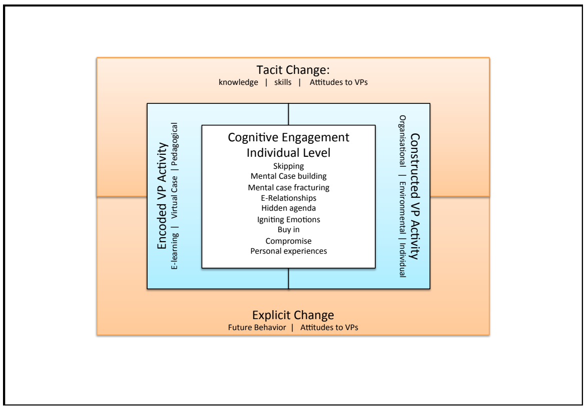
appendix 3 The iterative process to abstracting a model from the VP data



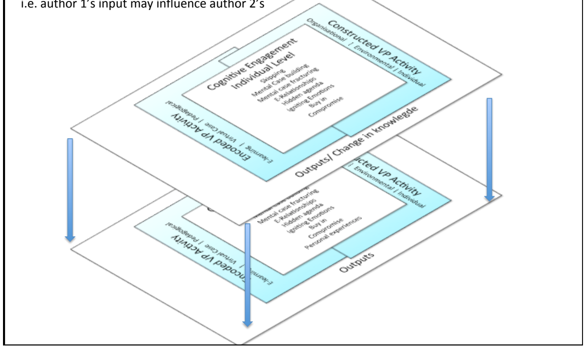


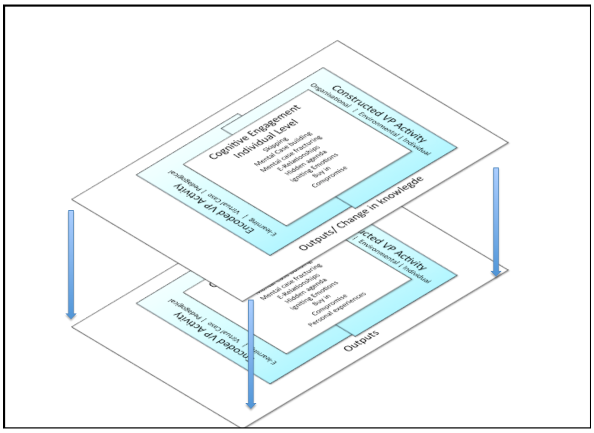


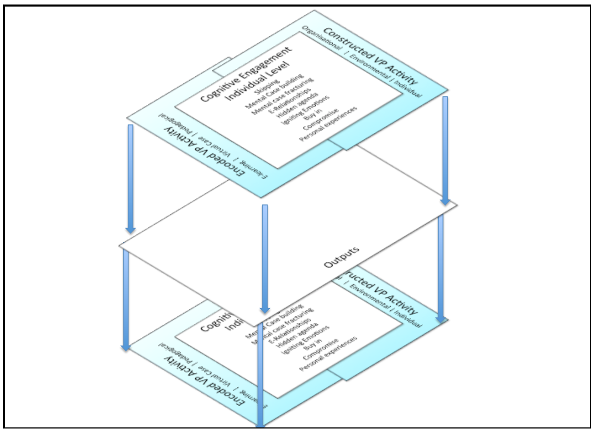


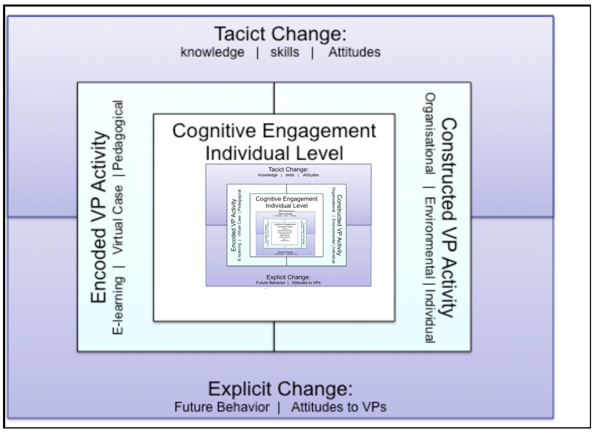


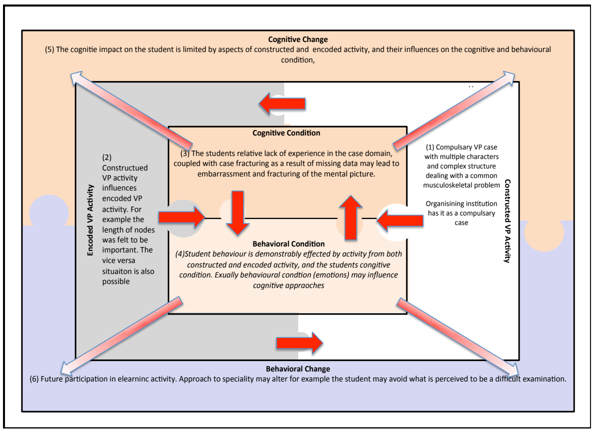
Virtual Stacks rather than fractals
Quite nice way of constructing Single VP experiences that are linked
i.e. case 1 may influence case 2
i.e. author 1's input may influence author 2's

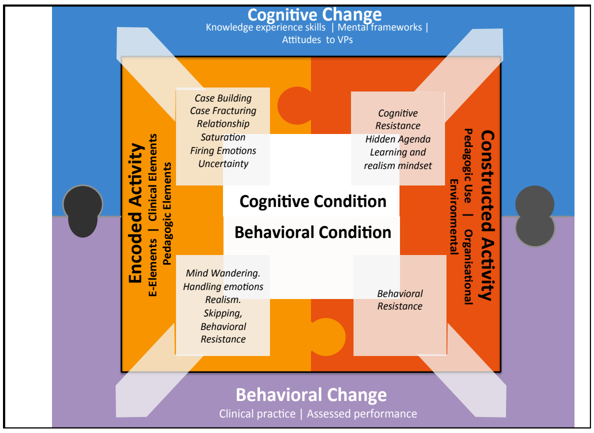


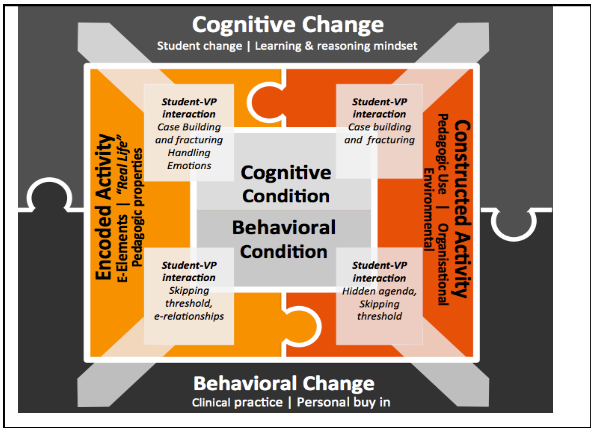

Cognitive Engagement
Individual Level
Constructed VP Activity
Encoded VP Activity
Outputs/ Change in knowledge
Outputs


Cognitive Engagement
Individual Level
Constructed VP Activity
Encoded VP Activity
Outputs

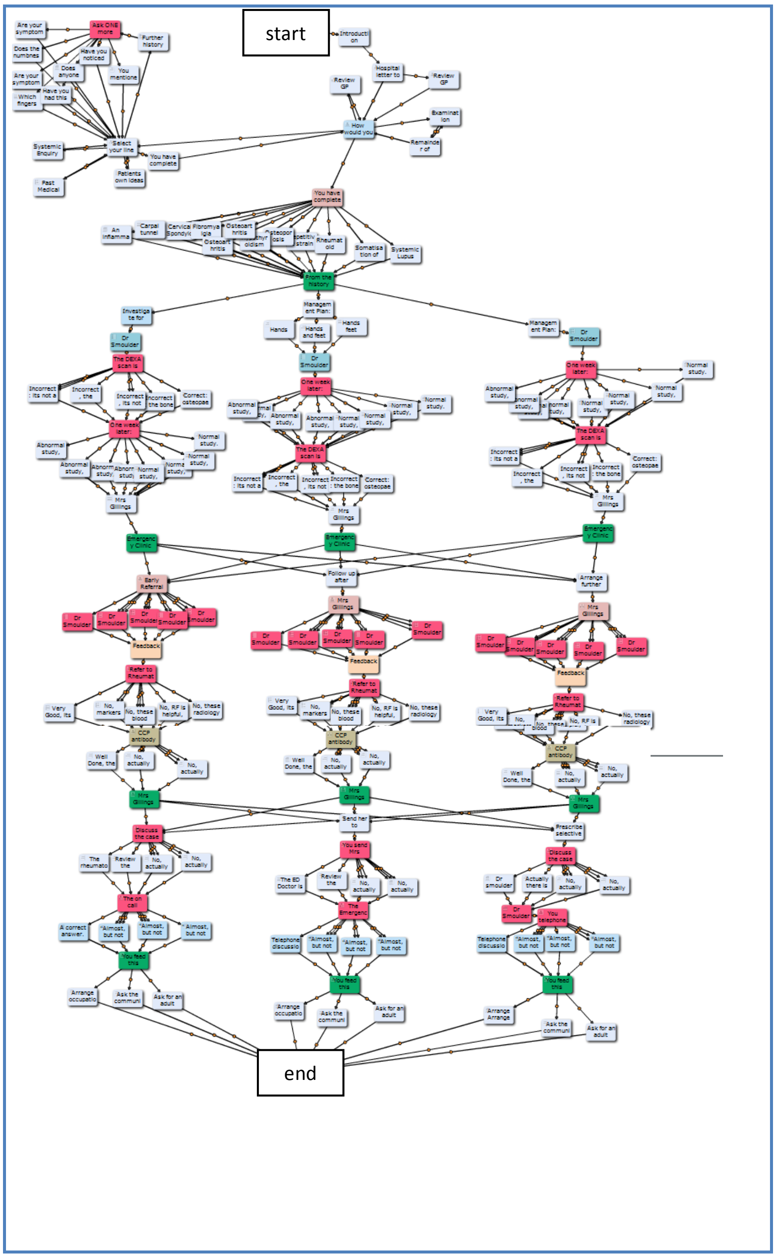

Tacict Change:
knowledge | skills | Attitudes
Cognitive Engagement
Individual Level
Encoded VP Activity
E-learning | Virtual Case | Pedagogical
Constructed VP Activity
Organisational | Environmental | Individual
Explicit Change:
Future Behavior | Attitudes to VPs


Cognitive Change
(5) The cognitie impact on the student is limited by aspects of constructed and encoded activity, and their influences on the cognitive and behavioural condition,
Cognitive Condition
(3) The students relative lack of experience in the case domain, coupled with case fracturing as a result of missing data may lead to embarrassment and fracturing of the mental picture.
Encoded VP Activity
(2) Constructued VP activity influences encoded VP activity. For example the length of nodes was felt to be important. The vice versa situaiton is also possible
(1) Compulsary VP case with multiple characters and complex structure dealing with a common musculoskeletal problem
Organisining institution has it as a compulsary case
Constructed VP Activity
Behavioral Condition
(4)Student behaviour is demonstrably effected by activity from both constructed and encoded activity, and the students congitive condition. Exually behavioural condtion (emotions) may influence cognitive appraoches
Behavioral Change
(6) Future participation in elearninc activity. Approach to speciality may alter for example the student may avoid what is perceived to be a difficult examination.


Cognitive Change
Knowledge experience skills | Mental frameworks | Attitudes to VPs
Case Building
Case Fracturing
Relationship
Saturation
Firing Emotions
Uncertainty
Cognitive
Resistance
Hidden Agenda
Learning and
realism mindset
Cognitive Condition
Behavioral Condition
Encoded Activity
E-Elements | Clinical Elements
Pedagogic Elements
Constructed Activity
Pedagogic Use | Organisational
Environmental
Mind Wandering.
Handling emotions
Realism.
Skipping,
Behavioral
Resistance
Behavioral
Resistance
Behavioral Change
Clinical practice | Assessed performance


Cognitive Change
Student change | Learning & reasoning mindset
Student-VP
interaction
Case Building
and fracturing
Handling
Emotions
Student-VP
interaction
Case building
and fracturing
Cognitive
Condition
Behavioral
Condition
Encoded Activity
E-Elements | "Real Life"
Pedagogic properties
Constructed Activity
Pedagogic Use | Organisational
Environmental
Student-VP
interaction
Skipping
threshold,
e-relationships
Student-VP
interaction
Hidden agenda,
Skipping
threshold
Behavioral Change
Clinical practice | Personal buy in

appendix 4 Reproduction of case authoring node sequence showing linking of nodes in a branching case used in this research.



Note: this is a representation from the computer authoring process, and is intended to show links between nodes (shaded boxes) from the top of the case (start) to the end (end). The connecting lines represent potential routes through the cases from start to finish.

appendix 5 Example Participant Information sheet, University of Birmingham, main quantitative study

V3.0

Participant Information sheet: **Virtual Patients in Rheumatology and Orthopaedics (ViPRO)**
Researcher: Dr James Bateman **Supervisor: Dr David Davies, Warwick Medical School**

Dear Medical Student,

I would like to invite you to take part in an educational research study funded by Arthritis Research UK taking place at Keele. Before you decide, you need to understand what the research is about, why the research is being done and what it would involve for you. Please take time to read the following information carefully. Feel free to talk with other students about the study. Please ask if there is anything that is not clear or if you would like more information about the study before you make your decision. Please take your time in making the decision to take part. Some questions below will help inform your decision. There will be a presentation about the research on the day of the study.

What is the purpose of the study?
We are interested in finding out the best way to design online virtual patients. We would like to study how the way the patients are designed influences how useful they are, to help develop future virtual patients as open access high quality resources.

Why have I been invited?
We are inviting all the medical students from Keelerotating through the musculoskeletal block to take part.

Do I have to take part?
Your participation is completely voluntary, and you can withdraw at any stage.

What will taking part involve for me?

- You will have to give your consent to take part if you wish to complete the cases
- You will 'log in' to the virtual patient cases using a link from the module web pages using your email address and a password that will be sent to you.
- You will briefly be asked some questions about yourself for the study
- You will then work through the fictional virtual patient cases, which will be online
- There are four cases, they are intended to be educational and are not a 'test'. You cannot pass or fail the cases; they are intended to help you learn.
- You will get feedback as you work through the cases to help you learn about the conditions. The feedback may be given in different for different cases
- All students taking part will compete the same core topics, but the exact design of the cases will vary from student to student (this is the main point of the research).
- You will be asked to fill in a short evaluation of the virtual patients.
- The virtual patients will **not** form part of any formal assessment, and your scores will be anonymised.

What will I have to do?
If you are interested and willing to take part we will ask you to give your consent, which is taken separately.

What are the possible benefits of taking part?
The work will give you the opportunity to work through some musculoskeletal medicine cases, and may prompt other questions you can discuss with doctors as you sit the block. In other studies students have enjoyed using virtual patients, and found them helpful.

Are there any risks in taking part?
The cases will take time to compete, but hopefully this will be offset by the educational benefit you get from them. The study has been reviewed by the Warwick Biomedical Ethics Committee (BREC), and the Keele School of Medicine Ethics Committee (SOMEC).

Will any other data about me be collected?
No other data is being collected from Birmingham University.

How will I find out about the results of the study?
Limited Results of the study will be published online at www.arthritisresearchuk.org, along with details of how to access the complete results once published, hopefully within 12 months of you completing the cases

What if I have a complaint?
As per the University's policy, any complaints can be made to the Deputy Registrar for governance. The contact details are: Nicola Owen, Deputy Registrar, Deputy Registrar's Office University of Warwick, Coventry CV4 8UW
T: 024 7652 2785 E: Nicola.Owen@warwick.ac.uk

What if I have any other questions?
Please ask the researcher in person, who will be available at a number of the academic in-days, or via email at james.bateman@warwick.ac.uk. The project supervisor can be contacted at david.davies@warwick.ac.uk.

Thank you for reading this information sheet,
Dr James Bateman

Appendix 6 Inter-item correlation matrix showing Pearson's R for the EViP items. The four subdomains are shaded, Cronbach's alpha is 0.849 for these eleven items.

	Authenticity		Professionalism				Coaching			Learning	
	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11
Q1	1	0.57	0.517	0.486	0.29	0.492	0.425	0.536	0.461	0.503	0.486
Q2	0.57	1	0.474	0.403	0.388	0.437	0.356	0.45	0.365	0.46	0.45
Q3	0.517	0.474	1	0.525	0.41	0.51	0.355	0.463	0.387	0.416	0.492
Q4	0.486	0.403	0.525	1	0.398	0.51	0.389	0.49	0.417	0.414	0.555
Q5	0.29	0.388	0.41	0.398	1	0.425	0.203	0.273	0.227	0.336	0.404
Q6	0.492	0.437	0.51	0.51	0.425	1	0.422	0.505	0.43	0.467	0.686
Q7	0.425	0.356	0.355	0.389	0.203	0.422	1	0.608	0.408	0.474	0.497
Q8	0.536	0.45	0.463	0.49	0.273	0.505	0.608	1	0.566	0.499	0.571
Q9	0.461	0.365	0.387	0.417	0.227	0.43	0.408	0.566	1	0.416	0.498
Q10	0.503	0.46	0.416	0.414	0.336	0.467	0.474	0.499	0.416	1	0.504
Q11	0.486	0.45	0.492	0.555	0.404	0.686	0.497	0.571	0.498	0.504	1

Appendix 7 Examples of research dissemination

The cases area available from go.warwick.ac.uk/msk a.) links to cases and social media support; b.) online supporting information; c.) student videos d.) QR code (quick response barcode) for conference presentations e.) example of social media engagement example via Twitter® from other users; f.) example of traditional print media presenting a one-page research digest

Appendix 8 National and International research poster and oral presentations (O/P).

Organised, Venue, and Date	Type:	Title (* denotes prize winner)
AMEE Conference. Prague, Czech Republic. August 2013.	Poster	Virtual Patients: 'This way' for evidence based, accessible, open-access resources
Association for the Study of Medical Education (ASME) Annual Scientific Conference, Edinburgh, June 2013.	Poster	Can open-access virtual patients be integrated into diverse UK medical schools? Results from a multi-centre study.
Association for the Study of Medical Education (ASME) Annual Scientific Conference, Edinburgh, June 2013.	O/P	How should virtual patients be designed for medical undergraduates? A multi-centre, randomised factorial study
Birmingham Education Conference, University of Birmingham, May 2013	O/P	Navigating the use and development of 'Virtual Patients': lessons and resources
Trainees for the Association for the Study of Medical Education 1st National Conference, Warwick, UK April 2013	O/P	Invited Speaker: 15 mistakes made in medical education research.
Association for the Study of Medical Education Europe (AMEE). Annual Conference, Lyon, France, August 2012	O/P	Researching Virtual Patients: A grounded theory study.
Warwick Medical School Annual Conference, Warwick, July 2013	Poster	Making a case- Virtual Patients*
British Society of Rheumatology Conference Glasgow SECC. May 2012.	O/P	Designing Virtual Patients for Musculoskeletal education: A grounded theory qualitative study.
British Society of Rheumatology Conference, May 2012. Glasgow SECC.	Poster	Open access musculoskeletal online education- virtual patients heading the call from the joints.
Birmingham Education Conference.' Advancing Quality in Education'. Birmingham University. April 2012.	O/P	Advancing the quality of education using virtual patients: A randomised controlled multi-centre experimental study
Higher Education Academy Health Sciences Annual Conference, Nottingham. May 2011.	O/P	Virtual Patients: Technology enhancing learning?
Association for the Study Of Medical Education (ASME). Brighton, June 2012. UK.	O/P	The impact of different virtual patient designs: a qualitative grounded theory focus group study.
Association for the Study Of Medical Education (ASME). Brighton, June 2012. UK.	O/P	Designing virtual patients to teach clinical reasoning: a randomised controlled multi-centre factorial study.
Association for the Study Of Medical Education (ASME). Brighton, June 2012. UK	O/P	The impact of different virtual patient designs: a qualitative grounded theory focus group study.
Warwick Annual Research Conference. 2012	Poster	A Year of Research
2nd International Digital Health Conference, Warwick, September 2011.	Poster	Virtual Patients: Researching evidence based designs to improve healthcare education and patient management*
AMEE Conference. Vienna, Austria. August 2011.	Poster	Using 'virtual patients' to teach medical undergraduate and postgraduate students: a qualitative study of different design properties
Arthritis Research UK Musculoskeletal Educators Conference, Brighton, 2011	O/P	Virtual Patients, the story so far.
Arthritis Research UK Annual Fellows Meeting. March 2011	Poster	Virtual Patients. Where are we now?

*Denotes poster prize winner

Appendix 9 Publications from this research

I have included the full text of four papers published in this appendix.

Appendix 8 a.)

Bateman, J., Allen, M., Samani, D., Kidd, J. & Davies, D. 2013. Virtual patient design: Exploring what works and why. A grounded theory study. *Medical Education,* 47, 595-606.

Appendix 8 b.)

Bateman, J., Allen, M. E., Kidd, J., Parsons, N. & Davies, D. 2012c. Virtual patients design and its effect on clinical reasoning and student experience: A protocol for a randomised factorial multi-centre study. *BMC Medical Education,* 12, 62.

Appendix 8 c.)

Bateman, J. & Davies, D. 2011. Virtual patients: Are we in a new era? *Academic Medicine : Journal of the Association of American Medical Colleges,* 86, 151; author reply 151.

Appendix 8 d.)

Bateman, J., Davies, D. & Allen, M. 2012d. Mind wandering has an impact in electronic teaching cases. *Medical Education,* 46, 235.

Other papers published as part of the context of this research project.

Bateman, J., Allen, M. E., Kidd, J., Parsons, N. & Davies, D. 2013. Virtual patients for rheumatology education: Preliminrary results from a multi-centre study. . *Rheumatology,* 52**,** i112.

Bateman, J., Allen, M. & Davies, D. 2012. Open access muscluoskeletal online education: Virtual patients are leading the way. *Rheumatology,* 52**,** iii157.

Bateman, J., Allen, M., Samani, D. & Davies, D. 2012b. Researching 'virtual patients': A grounded theory study. *Rheumatology,* 52**,** i44-45

Bateman, J., Hariman, C. & Nassrally, M. 2012. Virtual patients can be used to teach clinical reasoning. *The clinical teacher,* 9**,** 133-4.

Bateman, J., Francis, R. & Thistlethwaite, J. 2011. Medical student burnout and professionalism. *JAMA : the journal of the American Medical Association,* 305**,** 37-8; author reply 38.

Bateman, J. & Gull, D. 2011. Structural variations in attention-deficit hyperactivity disorder. *Lancet,* 377**,** 378-9.

Appendix 8 a.)

Bateman, J., Allen, M., Samani, D., Kidd, J. & Davies, D. 2013. Virtual patient design: Exploring what works and why. A grounded theory study. *Medical Education,* 47, 595-606.

virtual patients

Virtual patient design: exploring what works and why. A grounded theory study

James Bateman,[1,2] Maggie Allen,[2] Dipti Samani,[2] Jane Kidd[1] & David Davies[1]

OBJECTIVES Virtual patients (VPs) are online representations of clinical cases used in medical education. Widely adopted, they are well placed to teach clinical reasoning skills. International technology standards mean VPs can be created, shared and repurposed between institutions. A systematic review has highlighted the lack of evidence to support which of the numerous VP designs may be effective, and why. We set out to research the influence of VP design on medical undergraduates.

METHODS This is a grounded theory study into the influence of VP design on undergraduate medical students. Following a review of the literature and publicly available VP cases, we identified important design properties. We integrated them into two substantial VPs produced for this research. Using purposeful iterative sampling, 46 medical undergraduates were recruited to participate in six focus groups. Participants completed both VPs, an evaluation and a 1-hour focus group discussion. These were digitally recorded, transcribed and analysed using grounded theory, supported by computer-assisted analysis. Following open, axial and selective coding, we produced a theoretical model describing how students learn from VPs.

RESULTS We identified a central core phenomenon designated 'learning from the VP'. This had four categories: *VP Construction*; *External Preconditions*; *Student–VP Interaction*, and *Consequences*. From these, we constructed a three-layer model describing the interactions of students with VPs. The inner layer consists of the student's cognitive and behavioural preconditions prior to sitting a case. The middle layer considers the VP as an 'encoded object', an e-learning artefact and as a 'constructed activity', with associated pedagogic and organisational elements. The outer layer describes cognitive and behavioural change.

CONCLUSIONS This is the first grounded theory study to explore VP design. This original research has produced a model which enhances understanding of how and why the delivery and design of VPs influence learning. The model may be of practical use to authors, institutions and researchers.

Medical Education 2013: ***47****: 595–606*
doi: 10.1111/medu.12151

Discuss ideas arising from the article at www.meduedu.com 'discuss'

[1]Department of Medical Education Research and Development, Warwick Medical School, University of Warwick, Warwick, UK
[2]Medical Education Department, University Hospitals Coventry and Warwickshire National Health Service (NHS) Trust, Coventry, UK

Correspondence: Dr James Bateman, Department of Medical Education Research and Development, Warwick Medical School, University of Warwick, Warwick CV4 7AL, UK. Tel: 00 44 2476 574 880; E-mail: james.bateman@warwick.ac.uk

INTRODUCTION

Virtual patients (VPs) are computerised representations of realistic clinical cases.[1] It has been suggested that VPs are best applied to teach clinical reasoning skills.[2] Analogous to that on other web-based educational interventions, much of the focus on VP adoption has concerned 'if' rather than 'how' they should be designed and used.[3,4] Recent advances in technology and internationally adopted technical standards have potentially changed the definition of what a VP is as VPs can now be shared, edited and repurposed between institutions.[5,6] Virtual patients have the potential to deliver education to large numbers of students at a relatively low cost, which will be important in addressing the challenges that will face medical education in coming decades.[7,8] Existing principles for multimedia instructional design have been established[9,10] and studied in different instructional formats,[11] including preliminary work in VPs. For example, researchers have compared the delivery of text alone against that of text and supporting images.[12] Other general instructional design principles, such as those implied by cognitive load theory, are valid for some formats, including lecturing and e-learning.[13] However, such theory cannot be applied logically to VPs because they are often used to teach realistic scenarios in which the cognitive load is intentionally heavy in order to replicate that in the clinical scenario.

Given the wide range of VP design typologies, educational theory alone has limited ability to predict their success and the interplay among them.[14] The optimal design of VPs has therefore emerged as the subject of an important research question in the literature and has not been adequately addressed.[6] Qualitative methodologies including grounded theory provide opportunities to analyse, identify and explore experiences, and thus understand how and why different VP design properties support learning.

Objectives

This study has a principal aim and a subsidiary aim. The principle aim is to identify and explain how the design properties thought to be important in VPs influence medical student interaction with VPs. The subsidiary aim is to gain insight into which of the numerous VP design properties would be particularly relevant to study in future research.

METHODS

Study design

This is a grounded theory focus group study in which medical undergraduates from one medical school evaluate two substantial VP cases authored specifically for this research. We use one school of grounded theory, the 'classic grounded theory' approach proposed by Juliet Corbin and Anselm Strauss,[15,16] which uses iterative sampling, conceptual memoing, and the simultaneous collection and analysis of data. We consider our epistemological stance as analogous to that of Corbin.[16] We acknowledge the positivist origins of grounded theory, whilst recognising and valuing constructivism and reflexivity in the context of our pragmatist theoretical orientation and training.[16] The purpose of using grounded theory is to develop a theoretical explanation of 'what is going on', which, in this instance, refers to how and why design properties influence the effectiveness of VPs, by evaluating participant accounts and descriptions of their experiences. Our research protocol and research materials were subject to external peer review and were subsequently granted institutional ethics committee approval.

Virtual patient authoring

We authored two 30-minute VP cases, each of which intentionally included a number of design variables. We selected design features within the research group by consensus following a literature review of VPs and a review of publicly accessible VP cases. We chose only design features compatible with international VP interoperability standards produced by Medbiquitous.[17] We used the computer software DecisionSim Version 2.0 (Decision Simulation LLC, Pittsburgh, PA, USA) to author and deliver the cases following a technology appraisal of authoring tools. A detailed description and pictorial representation of our VP design variables can be found in supplementary material Table S1 and Fig. S1 (online). These variables included: branching and linear case narratives; freedom of navigation through the case; visible scoring systems; different question types such as multiple-choice questions; key feature problems, and Bayesian reasoning. Different learning strategies included: worked examples; listing of differential diagnoses; explicit identification of supporting information for and against particular diagnoses; information presentation techniques such as authentic clinical letters or salient results; post-case resources; data-gathering techniques (menu-driven history taking), and the provision of different feedback mechanisms with explicit and tacit feedback. These variables have been explicitly highlighted as important in both the VP and simulation literature.[2,6,14,18,19]

We integrated these designs into two substantial 30-minute VP cases focusing on core musculoskeletal medicine disease presentation. The cases were authored specifically for this research by an author experienced in producing several cases for a previous large VP research project.[20] We piloted the cases with two hospital consultant doctors, two general practitioners and two doctors in medical specialist training. This allowed us to correct grammatical, presentation and technical errors, judge the duration of the cases, and check item difficulty.

Participants and sampling

Participants were students in Years 2 and 4 at a graduate-entry medical school, which runs a 4-year programme. All were on placements at a university teaching hospital, which utilises VPs in some teaching modules. We used an iterative purposeful sampling technique to recruit volunteers.[21,22] Recruitment was voluntary; students were told the subject was musculoskeletal education research, but were blind to the use of e-learning. Students were given participant information sheets and asked to sign informed consent papers. There was no financial incentive to participate. We authored and piloted a funnelled questioning route using established focus group methodology, which we had piloted for acceptability, usability and question clarity.[23] This can be seen in Appendix S1 (online). We invited up to eight student participants to each focus group. We conducted six focus groups with a total of 46 participants. We used saturation sampling to decide when to terminate sampling as part of the constant comparative data collection and analysis.[21] As no new themes emerged during the sixth focus group, data collection was halted.

Data collection and analysis

Participating students completed the two VPs using a unique online identifier; after each case, students completed a self-report VP evaluation using an established VP evaluation instrument.[24] Data logs of student decisions and evaluations were recorded as additional sources of data to help inform the grounded theory analysis. They were explicitly not included in the protocol as subject to detailed statistical analysis for reasons of study design, sample size, power and sampling techniques. Following the evaluation, students took part in a 1-hour focus group discussion conducted according to an established methodology, with a pre-planned written funnelled questioning route. We used an experienced focus group facilitator (JB) and moderator (DS), who made field notes during the sessions. Focus groups were digitally recorded. The principal researcher transcribed, re-read and coded the interviews to develop a preliminary open coding structure, which was reviewed by other members of the research team. In line with the grounded theory approach, data analysis proceeded at the end of the focus group, prior to participant sampling for the next focus group. We used the computer-assisted qualitative data analysis software (CAQDAS) NVivo Version 9.0 (QSR International Pty Ltd, Doncaster, Vic, Australia) to facilitate analysis, interrogate the dataset, memo and provide an audit trail within the research team.[25] The research team included a medical educationalist, a university hospital head of medical education and a doctoral student. In our initial open coding we used a line-by-line approach, labelling phenomena using descriptive codes, and direct quotations from students. The latter are known by convention as *in vivo* codes.[26] We then proceeded to 'axial coding', a grounded theory method whereby codes are connected and grouped by virtue of common properties and dimensions.[15] We used scheduled regular research group meetings to describe, refine, review and seek alternative explanations for phenomena observed. Emergent themes guided sampling. For example, as student clinical experience emerged as an important theme, we purposefully sampled more junior students with little subject-specific experience in focus groups 5 and 6. Finally, in selective coding we identified a central core category which could be linked to all other categories using a paradigm described by the school of grounded theory.[15] Through a series of iterations and discussions within both the research group and institution educational board meetings, we then abstracted and verified a pictorial model describing how an individual student learns from a VP.[27]

RESULTS

Of the 48 students who volunteered to participate, 46 attended (19 males and 27 females). Of these, 31 students were in Year 4 and 15 were in Year 2. All students completed the VPs, evaluations and focus group discussion.

From the analysis theory we identified a core phenomenon, 'learning from the VP', to which four main categories related. As might be expected, these categories included *VP Construction*, which refers to material that an author integrates into a case; *External Preconditions*, which refers to student- and institution-centred factors; *Student–VP Interaction*, which refers to the fluctuating interface between the student and the VP, and *VP Consequences*, which refers to the after-effects on the student, which are co-dependent on the three prior categories. Examples of each category are given in Tables 1–3.

VP Construction

The first category to emerge from the data was *VP Construction*, which refers to the information encoded into a VP case file by an author. Three sub-categories emerged from the data: *clinical properties*, the clinical elements interwoven into a case; *pedagogic properties*, the educational characteristics of the e-learning case, and *electronic properties*, which include software, electronic usability, data presentation and the student–computer interface. These are defined individually alongside their component codes with examples in Table 1.

External Preconditions

The second category, of *External Preconditions*, emerged as the baseline characteristics outside the control of the student or authors that influence interaction with the VP. It has three sub-categories. *Student preconditions* describes how a student's clinical and educational experiences, attitudes, knowledge, skills and, in particular, negative experiences with e-learning, play a role in influencing that student's approach to VPs. The sub-category *organisational elements* relates to an institution's approach and policies that relate to curriculum, assessment strategy, learning environment, student appraisal and teaching materials. For example, the component 'Environmental factors' describes not only the location, but local factors such as the proximity of students to one another when sitting cases, and the computer hardware used to realise the authored cases. Each of these preconditions appears to influence how a student engages with a VP case. These sub-categories and component codes are defined with examples in Table 1.

Student–VP Interaction

The third category, *Student–VP Interaction*, is shown in Table 2 and describes the emergent interplay of elements between an individual student and the VP. This is specific to the individual student under the influence of the *External Preconditions* relating to the individual and the organisation, as well as the *VP Construction* (i.e. the first two categories described). This means individual instructional design elements produce different responses in different students. We identified six sub-categories of *Student–VP Interaction* that help to inform instructional design. The *skipping threshold* describes a point at which a student disengages with the case narrative, and the reasons why students skip through without reading or interacting with information presented. An example of how data triangulation has been used to support this phenomenon is shown in Table 4. Here, Student JC, who describes skipping (see quotation in Table 2), is shown to be actually skipping through content in comparison with colleagues and reviewers. We identified three broad reasons for this, which included a student's perspective of 'efficiency', the perceived credibility of the case, and the style of learning material. *Mental case building* and *mental case fracturing* describe how cognitive representations of cases are enhanced or sabotaged. *e-Relationships* describes the formation and evolution of relationships between the student and the people represented in the case. Components of *e-relationships* include a 'relationship threshold': as the number of characters in the VP rises above three, it becomes difficult for a student to maintain e-relationships with them. *Hidden agenda* reflects student behaviour in terms of the student's treatment of the VP as both a summative assessment with deliberate traps and triggers, and an evaluation subtext in which the student probes for author style, patterns and system vulnerability. *Handling emotions* describes how students deal with the 11 different emotions we identified, which include embarrassment and fear of working with 'virtual' colleagues. We present quotations from Student AP (Table 2) which show the embarrassment caused by making a mistake and seem to reflect the realism felt in the case. These emotions appeared to be reinforced by professional stereotypes influenced by a student's experience (*External Preconditions*, Table 1). For example, one view of the behaviour of hospital specialists was that they do not provide positive feedback, exemplified by the comment 'Consultants don't give out praise' (Student CB, focus group 2).

Table 1 The first two categories, 'VP construction' and 'External preconditions' of the central phenomenon 'learning from the VP'. In vivo *codes (direct quotes) are in italics with quotation marks*

Category 1. VP construction: The VP properties that are designed into a case
Clinical properties. Clinical properties authored into the virtual patient by an author
"**Real Life**": The clinical properties authored into a VP case and how they relate to actual clinical practice
Environment: Simulation of the clinical environment, for example GP having past health care records from a patient
Authenticity: The authenticity of the narrative and supporting educational materials in the case
"I like the way it's based on the way we've been taught so far… you start with the history and you take a detailed history, and I like that it actually gave you the option of collecting that history from that patient. … it still followed the steps that you would take in a normal situation which is getting a clear history, a systems review included of a patient and a condition… definitely something that applies to real life and definitely something that would be useful." EA, FG6, Year 2 student
Scope and content: The extent to which health care domains are explored by the case, such as clinical knowledge, professionalism, clinical reasoning, local health care policy, and health service structure
Pathway Flux: How the flow of clinical and other information is presented between the student and the VP
Channels and dams: The degree of freedom given to the student over their actions, progression and the narrative in the case
Evolution-Evaluation: The extent to which data and information is presented, reviewed and evaluated as the case progresses
Clinical Inertia: How case progression is resisted by the quantity, quality, completeness and relevance of pathways, data and activities that contribute to cognitive load, realism and difficulty
"the referral letter was good and bad, good because it's probably what we'd get, and bad because it was a bad referral letter… one of the questions was what is pertinent to this referral letter… and it had duration of symptoms, and you don't know how long its been going on for…" SR, FG3, Year 4 student
Pedagogic properties: Teaching elements of VPs integrated into VP cases
Feedback: How feedback is delivered to the students as they complete a VP case
Format effects: Implications of different formats such as a letter, or a phone call, at different times through the case
Tailoring: Extent to which student feedback is individualised, including comparisons with peer performance
Prompting reasoning: Approaches that explicitly drive structured clinical reasoning
"It was good to kind of think about the differentials… I do think the lack of knowledge was an influential factor, but it did help me question why is it that I'm including this one, and why is it that I'm including that one, I looked back to the history… you come across important factors… is that a long term condition, or is this acute… rule things out… I thought it was really good." RR, FG5
Decision Flux: How decisions contribute to freedom to make decisions both correct and incorrect, and experience consequences of them.
Consequence effect: Extent to which students feel their decisions impact further down the case narrative.
Limits and Forcing: Being forced to undertake a particular action, decision, cognitive process or clinical experience irrespective of the apparent choices given
"I quite liked the way that sometimes they got you to pick only three questions, which kind of got you maybe to think rather than ask just random questions. Think where your thoughts were going and what questions were important" CD. FG6
E-properties: Electronic properties used provoking comment and outside of normal expectations for electronic interfaces
E-Signposting: The helpful effect of signposting students using images of locations and particularly patients
"I really liked on the first case the pictures. I know, I know it was just random adjudicators, but it kind of made you smile and if you've got that kind of visual stimulation, oh that's the GP OK, it kind of motivates you…" AR, FG1
E-inertia: Electronic properties authored into cases which produce slow or hinder a student interacting with a case.
Non e-tasks: The use of items that don't require actions by the student for example summarising elements on paper
Software limits: Desired software features from students, not present, which limit interaction.
"***Scroll scroll scroll***"*: Impact of multimedia design including text format, length, steps, and image representation*
"I think some of the pages were quite wordy, maybe it can be broken down into two instead of one, and squeezing all of the information into one page, it just gives me a headache" SS, FG2
E-Error: Electronic error as students sit a case, the cause of which may or may not be under the control of the case author
Category 2. External preconditions: factors, external to the VP construction which influences its utility
Student centred: Student centred factors that influences the utility of the VP case
Electronic Prejudice: How both prior positive and negative e-learning experiences prejudice the approach to a VP
Global experience: The global knowledge and skills a student possesses about medical problems and health care systems
Student Goals: The students personal goals for the learning activity, for example relating to assessment or professional development.
"I can see why its on there, because for finals, they are going to say, "what are your differentials, summarise the case in a sentence" AR, FG1, Year 4 student
Organisation elements: An organisation's educational aims, curriculum, assessment and evaluation of students
Institution fingerprint: The organisational style and expectations of students when approaching educational/assessment
"I don't know how that reflects on our teaching… we're quite often lead to a single best answer" KG, FG1, Year 4 student
Assessed curriculum: Factors that relate to the pedagogic, assessment and curricular approaches of an educational institution
Environmental elements: Local factors (location, computer hardware) that influence how a student interacts with a VP case
"I know we had an hour, and at the end people were leaving so I felt, sort of, hurry up." DE, FG3, Year 4 student

GP = general practitioner; FG = focus group

Consequences

The fourth and final category is *Consequences*, which describes the result of the student–VP interaction (Table 3). *Consequences* has two sub-categories: *student change*, which refers to the learning that has occurred, and *preferences and buy-in*, which describes student attitudes to the VP completed and to future VPs. Students describe a unique 'individualised learning experience' based on their student–VP interactions shaped by VP construction and preconditions. In alluding to the 'learning–realism trade-off', we describe ideal VP characteristics that seemingly cannot both be addressed. An example in Table 3 describes the branching VP case as being more realistic, but as providing an inferior learning experience.

The model: VP implementation

The model we produced (Fig. 1) is based on and grounded in the data. This model describes and predicts the impact of different design-, institution- and student-related factors on the critical engagement between a student and the VP, the *Student–VP Interaction*. Ultimately, the model predicts the consequences of an individual sitting a VP, co-dependent on the prior factors described. The model has three layers. The inner layer is centred on the student, formed from the 'cognitive condition', and 'behavioural condition'. This describes the student's state prior to interacting with the VP, formed by prior attitudes, knowledge, experiences and skills. The middle layer describes VP delivery as comprising two elements. The first of these is as an 'encoded object' and refers to the VP's properties as an e-learning case. These include the electronic, pedagogic and clinical factors described. The second element describes VP delivery as a 'constructed activity', which encompasses how an institution delivers teaching using a VP case. 'Constructed activity' includes curriculum, pedagogic context and the environment in which VPs are delivered. Overlapping the inner two layers in our model is the *Student–VP Interaction*, which describes the critical behavioural interactions dependent on the student, and encoded and constructed activity. The outer layer represents the product of the interaction, learning, as 'cognitive change' and 'behavioural change'.

DISCUSSION

We have explored a number of design features in VP cases created to established technical standards in order to build theory in this important and under-researched area.[6] This original research has produced a model (Fig. 1) that explains how and why different VP design features influence learning. We describe a series of six new phenomena within the *Student–VP Interaction* that help to give a deeper understanding of how and why a student engages with and learns from a VP. Clinicians may or may not be aware of these interactions and the importance and impact of encoded and constructed activity. Thus, they may not be either consciously or unconsciously incorporating design variables to accommodate these factors.

The multiple design variables used to author the VPs are part of the encoded activity. Some design properties produced a mixture of positive and negative effects. These included: case pathway choices, such as branching versus linear case narratives; some structured approaches to clinical reasoning using the evidence available, and data quantity and presentation (such as the use of additional information, which added to 'clinical inertia', and scoring counters). Some elements intended to promote sound clinical reasoning appeared at times to frustrate students, such as prompting 'non e-tasks' and some case presentation 'e-elements', such as the scoring counter. As evidence emerges for these approaches in clinical medicine, careful consideration to their application can be informed by our model. This could include incorporating structured reasoning as an electronic task, and limiting the intrusion of these processes into the realism of a case.

For example, the scoring counter helped some students in 'learning and assessment focus', but produced negative behaviours including that of lowering the *skipping threshold* in some students seeking to obtain 'points on the board' (Table 2).

The model highlights a trade-off between certain design possibilities and student behaviour and the consequences of these for learning. Our model prompts authors to consider the opportunities and constraints of 'constructed activity', such as those imposed by technical facilities, curricular integration and participants, and the attributes of the VP audience, the 'student preconditions'. Each apparently desirable design feature, such as complex branching with realistic 'clinical inertia', has positive and negative effects on students. The model predicts that using a branched case design potentially makes cases more realistic, but is associated with increased

Table 2 Student–Virtual Patient (VP) Interaction, the third category from the central phenomenon 'learning from the VP'. In vivo codes (direct quotes) are in italics with quotation marks

Category 3. Student–VP Interaction: the interaction between a student and a VP as the student completes a case
Skipping threshold: a threshold above which negative behaviour patterns occur and interaction ceases to be constructive to learning
Efficient skipping: engagement is limited by a drive to efficiently pick out activities that are perceived to add value or be important
JC: 'Some of the pages had a lot of words on, and my eyes go, nah there's a lot to read there, and there's nothing to input, and I don't have to give anything, so therefore there's no need for me to read it because it wasn't about the case. That's just me being lazy...
AA: 'I think by nature we're all quite lazy, and we'll be like "Nah"'JC: 'Efficient' JC and AA, Year 4 students, FG2
Judged credibility: a constant appraisal of the usefulness, quality and interactivity of the VP to judge whether to continue with the case
Style: approaches in which the style of questioning promoted lack of engagement and skipping
'There were some of the questions where it was you selected one answer, and it wouldn't allow you to go through until you clicked the right one... At that point I'd given up and was just guessing, which obviously I would never do with a real patient'DG, Year 2 student, FG5
Mental case building: interactions which help the student to construct a mental representation of the case
Learning and assessment focus: added engagement resulting from perceived benefits in learning, future assessments or the workplace
'Thinking outside the list': strategies which encourage students to think outside a predetermined list of answers in the task
Pathway growth: the extent to which both decisions and branching pathways enhance students' experiences, and experiential learning
'I think I quite like the branching bit... because obviously in life there are a lot of different routes you can take and it doesn't necessarily mean one is... the best ... I think it's good to go a little bit off track'JC, Year 4 student, FG2
Mental case fracturing: interactions which impair the student's construction of a mental representation of the case
'Bogged down': role of ultimately irrelevant information either explicit (doctor suggesting behaviour for student) or tacit (information load)
Invisible elephant: the extent to which students see or do not see feedback that is integrated into the case narrative, but not explicitly labelled
'Loss of control': students' perceived loss of control in the case that may or not be related to branching structures
Pathway decay: decay in learning, which occurs as a result of losses in time or effort, uncertainty or motivation caused by being allowed to follow different routes
Contextual dissonance: factors which clash with previous case assumptions, such as discordant information
'There were some discrepancies... the age on the GP records is different to when you are given the first stem, there is a 9-year difference, she was born in 1970 in one and 1979, I don't know whether or not that's relevant'FH, Year 4 student, FG1
False expectations: students' false preconceptions about clinical scenarios and professional duties that are detrimental
KG: 'That's the sort of thing you're going to get asked in an exam. Clinically I don't think it's that relevant. I think being able to say, is it likely or is it unlikely 'FH: 'I don't think we should have to work it out for ourselves...'KG and FH, Year 4 students, FG1 [discussing Bayesian reasoning with immunology laboratory tests]
Points on the board: the student's primary focus becomes the assessment and scoring employed during the case
e-Failure: a technical failure, the origins of which may be the student, author, VP software, or IT software, hardware and infrastructure
e-Relationships: how the students form relationships with electronic representations of patients and health care professionals
Stereotyping: stereotyping or making moral professional or personal judgements about case participants
FK: 'She was like "Oh, I have a new partner, I want to start a new family"'
AR: 'I was thinking, you've left it a bit late'FK and AR, Year 4 students, FG1
Relationship threshold: complexity of sustaining more than two relationships in a case (with supervisors, the patient, allied health professionals)
Hidden agenda: activity of deconstructing the VP and its components either naturally, out of curiosity or to improve performance
Assessment subtext: interpreting the case in the context of the institution or teacher assessment strategies
MB: 'And when you think of it as an exam, you start looking for a style, because everyone has a style in the way they will write a question and answer, and you're trying to link the two up rather than thinking...'JC: 'What actually should I do'JC, MB and JC, Year 4 students, FG2,
Case template subtext: the student devotes time to exploring real or perceived examiner VP design structure, for interest or to find patterns of assessment
'It felt to me like essentially you went off for a little tangent for a couple of windows and then it would drop you back onto a common pathway towards to the end of whatever you did' JW, Year 4 student, FG1
Handling emotion: students described the process of coping with different emotions during the case, 11 in total:
Fairness, Humour, Comfort zone, Uncertainty, Fear, Confidence, Denial, Embarrassment, Pressure, Fatigue and Distraction
'It seemed quite realistic to me, like kind of both embarrassing and reassuring at the same time. Even though it was simulated, I did feel a bit embarrassed when I was being slightly corrected when I hadn't decided to refer the patient... and so had... "Oh, really you should have referred that... but don't worry I've sent off the referral."... That's... giving you the feedback in a realistic way, how it probably would happen in real life, and I feel like I'm going to remember that a lot more because of that feeling of embarrassment'AP, Year 2 student, FG5

GP = general practitioner; FG = focus group

complexity and a higher frequency of and more types of 'e-inertia' and errors (Table 1). The resulting effects this appears to have on *Student–VP Interaction* in terms of the *skipping threshold*, 'cognitive dissonance' and 'case fracturing' (Table 2) are complex and different for each student. These cannot be abstracted from this model to a general recommendation supporting or opposing a design variable

Table 3 The fourth category from the central phenomenon 'learning from the VP', 'consequences'

Category 4. Consequences:. The results of a student engaging with an individual or series of VPs
Student change: The impact on knowledge and behaviours in students future practice
Real world reasoning: Incorporating processes taught in the VP that changes clinical practice and approach to patients
"I also liked the multiple choices question part, despite having 10 options for the blood results, what are the three most important ones.....It gets you into the mind-set of not ticking all of the boxes, which in theory you could probably do come August if you wanted to."
BN, FG4, Year 4 student, [NOTE: August refers to the month graduates begin work as a qualified doctor]
Addressing weakness: Highlighting areas of knowledge skills or behaviours that are weak, and addressing those areas
"I kind of guessed the first one, and then I realised actually you can work it out... so it highlights your weaknesses I suppose"
AR, FG1, year 4 student
Individualised experience: Unique user learning experiences which depends on domains one to three
"I have a different experience from you, again because I didn't look at the score...I stopped and I started doing the modified Schober's test... having that break where I didn't feel like I had to do anything, I was just learning."
AL, FG1, year four student
Personal 'Buy In' extent to which VP design influences current and future participation of VP cases
Learning-realism trade-off: An apparent trade off between learning and realism when faced with different design properties
"I think the question... is.... do you want learning or realism, because it was better to learn with the linear case, because obviously there's only one way to go with it... if you make the wrong decision...we're not going to learn what the right path necessarily is. Whereas its going to obviously be more realistic... so the second case was more realistic but the first one was a better learning experience."
MB, FG2, year 4 student
Future uptake: The approach to voluntary or compulsory cases in future training

such as 'branching', but can only help to provide context for that decision based on the other components of the model (the students, constructed activity).

We did find general consensus in favour of and against some design properties. 'e-Signposting' and the use of 'key feature problems' and Bayesian reasoning questions were helpful; the latter two represent existing validated measurements of clinical reasoning skills.[28] The item difficulty in these questions produced interesting phenomena such as that of 'perceived e-error'. Poorly received design properties included problems with the narrative ('relationship saturation'), information presentation ('scroll scroll scroll', an 'e-property) (Table 1), and feedback (the 'invisible elephant', a component of 'case fracturing') (Table 2).

This was an innovative research project in two key areas: the research methodology, and its transparent reporting. Firstly, our methodology utilised available computer-assisted technology to identify, describe and triangulate new findings. These data were used to help understand individual interactions with VPs, and neither used nor planned for quantitative analysis. To our knowledge, this is the first VP research study to explicitly construct VPs to established software standards, with the single purpose of researching important design properties such as branching cases. Our second-by-second individual student data logs and self-reported evaluations (Table 4) allow the triangulation and exploration of theoretical constructs. These have contributed to the descriptions of new and interesting phenomena in *Student–VP* Interaction, such as the *skipping threshold*, and link described with actual behaviour (Table 4), facilitating comparison with observed clinical practice. We consider this skipping to be in many ways analogous to mind wandering.[29] Secondly, our open transparent representation of research includes unprecedented detail in both the descriptions and schematics of the VPs used (Table S1 and Fig. S1) and the XML case files (details available from the authors), which have not been included in recent research into VPs.

This work builds on the existing literature on VPs, and emphasises that what an author encodes is only one of many variables contributing to educational impact. This model adds to research on design principles,[2, 6, 12] practical authoring advice,[30, 31] curricular integration,[32] the theoretical principles behind the VP,[33] the importance of the environment,[34] and the difficulties inherent in sharing and repurposing VPs.[35] Our work supports 10 general authoring recommendations produced from a thematic analysis of VPs in focus groups,[12] and provides a framework within which any characteristic, such as authenticity, can be considered within our model.

Table 4 An example of three sources data triangulation

(a) Shows Student JC who had described skipping content in the focus group to have actually skipped content using our data logs. JC spent three-seconds on examination findings, shorter than peers and two reviewers, but performed satisfactorily compared to peers on the case score

User	Seconds spent on window "Examination of Mrs Begum"	Case score
JC*	**3**	**13**
CB	40	18
JM	25	12
HD	33	16
AA	30	13
MB	52	12
AA	23	15
AM	33	12
Reviewer 1	23	N/A
Reviewer 2	15	N/A

(b) Shows an example of free text feedback, prior to participating in focus group

"I thought the use of lots of information from the history, test results and radiographs made it a very realistic case to work through."
WA, Electronic free text comment, prior to FG3, Year 4 student
"It was useful as I had to make the decision on what I would do, but if wrong I was redirected on the right course so I could learn from the case."
JM, Electronic free text comment, prior to FG2, Year 4 student

(c) Is a demonstration of CAQDAS supporting the choice of an 'in vivo' code, "real life"- the phrase occurs 65 times, and in all focus groups

Focus group	Year of training	Occurrence of phrase "real life"
1	4	7
2	4	11
3	4	4
4	4	25
5	2	13
6	2	5
Total		**65**

Limitations

Some design elements lay outside the scope of this investigation. We chose not to study video or audio files as design properties for reasons of cost, technological limitations for distribution, and lack of flexibility for updating and repurposing. We did not study natural language input or free-text questioning. Although equivalents have been used in some VP research,[12] they are unavailable in most open-source and some commercial VP players[31] and to our knowledge do not have an open technical standard equivalent.[17] We cannot exclude the possibility that the presentation effects of the software interface we used, DecisionSim, may have an impact. We did not investigate a number of areas, such as the use of VPs as part of group work,[6, 36] different curricular integration strategies,[24, 32] novel approaches to teaching clinical reasoning such as the 'think-aloud' approach,[37, 38] or address specific situations such as cognitive bias.[39] That said, most of these approaches describe either how an institution deploys a VP, and thus might fit into our model as 'constructed activity'; the others reflect elements of the encoded activity.

We have attempted to address as far as possible the problems associated with bias, reflexivity, observer effects and researcher preconceptions using a number of methods. When considering reflexivity, the impact of the researchers' position and perspectives, we acknowledge that we consider that VP design is important and this stance may potentially result in an inductive bias and influence the participants,

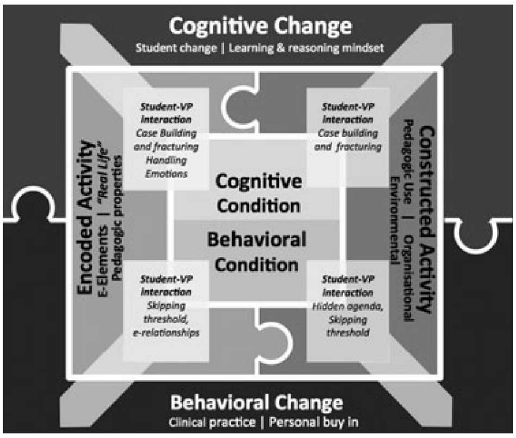


Figure 1 Virtual patient (VP) implementation model. It includes three layers in which *Student–VP Interaction* overlaps the inner two layers and describes how the different ways in which VPs are implemented can influence learning

analysis and conclusions. We argue that our subjection of the research protocol to an iterative process of both internal and external peer review has helped to prevent this, as have regular institution review board meetings at which analysis was presented.[15, 16, 40] Other grounded theorists would question the validity of many of these approaches and, for example, may not agree with our use of multiple data sources and triangulation.[16] All of these approaches would be supported by the school of grounded theory to which we subscribe.[15] Participants were part of a study in one institution, which raises the possibility of observer effects.[41] To minimise these, the research study was completed in a familiar learning environment. Our research may not be transferable to other institutions and health care professional groups. The extent to which different groups rely more or less on VP cases in their curricula may also influence this.[7] We have observed that our themes did recur across year groups, in students who had and had not previously used VPs. We did not seek 'respondent validation' for our theory; we are in agreement with the literature that proposes this represents another data source for analysis rather than a form of theory validation.[42, 43]

Practical implications for medical education

We have produced a model that has practical use for stakeholders who may be authoring or commissioning VPs. The model describes how VPs produce different learning experiences in a framework that incorporates students, authors, technical and software elements, institutions and environments. Developers can consider which design features and electronic presentation to adopt, the 'encoded activity', and how they should be delivered, the 'constructed activity'. In a consideration of these and 'student preconditions', which refers to, for example, how electronic prejudices are formed, and how they all influence types of *Student–VP Interaction*, the present framework may help to elucidate how a VP might be developed for any given topic. Researchers can also use the framework to consider how to plan and report new research to help inform VP adoption, repurposing and sharing, against a backdrop of challenges to the resourcing of education for medical and allied health professionals.[44] For example, a senior allied health professional teaching on a national Masters-level course could consider the impact of student preconditions and how existing VP cases might be adapted or repurposed (encoded activity) within the environment and resources (constructed activity) to achieve the objectives inherent in the use of VPs (cognitive and behavioural change).

We hope other educators, researchers and institutions will incorporate this model when developing and researching VPs, and call for researchers to follow transparent reporting of VP design properties. We hope this work will help to inform the interpretation of further research we are conducting into measuring the impact of different VP designs.[45]

Contributors: all of the authors made substantial contributions to the research design and study protocol. JB, MA and DD conceived the original study design. JB wrote the initial draft of the study protocol, which was revised and

approved by the other authors. DS assisted in facilitating the focus groups. JB conducted the main analysis supported by DD, MA and DS. JB drafted the initial write-up, which was revised with input from all authors. All authors approved the final manuscript for publication.
Acknowledgements: We wish to thank the Arthritis Research UK Education and Training Committee and Professor Jill Thistlethwaite, from the University of Queensland for review of the initial research protocol. We thank the students and faculty from Musculoskeletal and Medical Education at the host institution. We also wish to thank like to thank the Dr JB McGee from the University of Pittsburgh, and Decision Simulation, who produced the VP software, DecisionSim, for the study.
Funding: This study was funded via an education research fellowship by Arthritis Research UK, Grant 19330.
Conflicts of interests: none.
Ethical approval: this study was approved by the Warwick Medical School Institutional Ethics Review Board and the Biomedical Research Ethics Committee.

REFERENCES

1 Ellaway R, Poulton T, Fors U, McGee JB, Albright S. Building a virtual patient commons. *Med Teach* 2008;**30** (2):170–4.
2 Cook DA, Triola MM. Virtual patients: a critical literature review and proposed next steps. *Med Educ* 2009;**43** (4):303–11.
3 Cook DA, Levinson AJ, Garside S, Dupras DM, Erwin PJ, Montori VM. Internet-based learning in the health professions: a meta-analysis. *JAMA* 2008;**300** (10):1181–96.
4 Cook DA, Levinson AJ, Garside S, Dupras DM, Erwin PJ, Montori VM. Instructional design variations in Internet-based learning for health professions education: a systematic review and meta-analysis. *Acad Med* 2010;**85** (5):909–22.
5 Bateman J, Davies D. Virtual patients: are we in a new era? *Acad Med* 2011;**86**(2):151.
6 Cook DA, Erwin PJ, Triola MM. Computerised virtual patients in health professions education: a systematic review and meta-analysis. *Acad Med* 2010;**85** (10): 1589–602.
7 Huang G, Reynolds R, Candler C. Virtual patient simulation at US and Canadian medical schools. *Acad Med* 2007;**82** (5):446–51.
8 Frenk J, Chen L, Bhutta ZA *et al.* Health professionals for a new century: transforming education to strengthen health systems in an interdependent world. *Lancet* 2010;**376** (9756):1923–58.
9 Clark RC, Mayer RE. *E-learning and the Science of Instruction: Proven Guidelines for Consumers and Designers of Multimedia Learning,* 2nd edn. San Francisco, CA: Pfeiffer 2008;49–62.
10 Mayer RE, Griffith E, Jurkowitz IT, Rothman D. Increased interestingness of extraneous details in a multimedia science presentation leads to decreased learning. *J Exp Psychol Appl* 2008;**14** (4):329–39.
11 Issa N, Schuller M, Santacaterina S, Shapiro M, Wang E, Mayer RE, DaRosa DA. Applying multimedia design principles enhances learning in medical education. *Med Educ* 2011;**45** (8):818–26.
12 Huwendiek S, Reichert F, Bosse H-M, de Leng BA, van der Vleuten CPM, Haag M, Hoffman GF, Tönshoff B. Design principles for virtual patients: a focus group study among students. *Med Educ* 2009;**43** (6):580–8.
13 van Merrienboer JJG, Sweller J. Cognitive load theory in health professional education: design principles and strategies. *Med Educ* 2010;**44** (1):85–93.
14 Huwendiek S, de Leng BA, Zary N, Fischer MR, Ruiz JG, Ellaway R. Towards a typology of virtual patients. *Med Teach* 2009;**31** (8):743–8.
15 Corbin JM, Strauss AL. *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory,* 3rd edn. London: Sage Publications 2008;65–87.
16 Bryant A, Charmaz K. *The SAGE Handbook of Grounded Theory.* London: Sage Publications 2007;191–213.
17 Medibiquitous Virtual Patient Working Group. MedBiquitous Virtual Patient Player Specifications and Description Document V.1.0. 2010. http://www.medbiq.org/working_groups/virtual_patient/VirtualPatientPlayerSpecification.pdf. [Accessed 4 February 2013.]
18 Issenberg SB, McGaghie WC, Petrusa ER, Lee Gordon D, Scalese RJ. Features and uses of high-fidelity medical simulations that lead to effective learning: a BEME systematic review. *Med Teach* 2005;**27** (1):10–28.
19 Norman GR, Cook DA. Simulator training for recognition of murmurs: give a kid a hammer? *Chest* 2011;**139**(5):1257–8.
20 Smothers V, Ellaway R, Balasubramaniam C. eViP: sharing virtual patients across Europe. *AMIA Annu Symp Proc* 2008:**6**:1140.
21 Strauss AL, Corbin JM. *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory.* 2nd edn. London: Sage Publicatoins 1998;12–52.
22 Kennedy TJT, Lingard L. Making sense of grounded theory in medical education. *Med Educ* 2006:**101**:8.
23 Krueger RA, Casey MA. *Focus Groups: A Practical Guide for Applied Research,* 4th edn. Los Angeles, CA: Sage Publications 2009;25–62.
24 Huwendiek S, de Leng BA. Virtual patient design and curricular integration evaluation toolkit. *Med Educ* 2010;**44** (5):519.
25 Kelle U. *Computer-Aided Qualitative Data Analysis: Theory, Methods and Practice.* London; Thousand Oaks, CA: Sage Publications 2002;19–52.
26 Bourgeault IL, Dingwall R, De Vries RG. *The SAGE Handbook of Qualitative Methods in Health Research.* Los Angeles, CA: Sage Publications 2010;289–306.
27 Strauss AL, Corbin JM. *Basics of Qualitative Research: Grounded Theory Procedures and Techniques.* London: Sage publications 1990;8–28.

28 Page G, Bordage G, Allen T. Developing key-feature problems and examinations to assess clinical decision-making skills. *Acad Med* 1995;**70** (3):194–201.
29 Bateman J, Davies D, Allen M. Mind wandering has an impact in electronic teaching cases. *Med Educ* 2012;**46** (2):235.
30 Posel N, Fleiszer D, Shore BM. 12 Tips: guidelines for authoring virtual patient cases. *Med Teach* 2009;**31** (8):701–8.
31 Begg M. Virtual patients: practical advice for clinical authors using Labyrinth. *Clin Teach* 2010;**7** (3): 202–5.
32 Berman N, Fall LH, Smith S *et al.* Integration strategies for using virtual patients in clinical clerkships. *Acad Med* 2009;**84** (7):942–9.
33 Ellaway RH, Davies D. Design for learning: deconstructing virtual patient activities. *Med Teach* 2011;**33** (4):303–10.
34 Edelbring S, Dastmalchi M, Hult H, Lundberg IE, Dahlgren LO. Experiencing virtual patients in clinical learning: a phenomenological study. *Adv Health Sci Educ Theory Pract* 2011;**16** (3):331–45.
35 Fors UG, Muntean V, Botezatu M, Zary N. Cross-cultural use and development of virtual patients. *Med Teach* 2009;**31** (8):732–8.
36 Vash JH, Yunesian M, Shariati M, Keshvari A, Harirchi I. Virtual patients in undergraduate surgery education: a randomised controlled study. *ANZ J Surg* 2007;**2**:54–9.
37 Lee JE, Ryan-Wenger N. The 'Think Aloud' seminar for teaching clinical reasoning: a case study of a child with pharyngitis. *J Pediatr Health Care* 1997;**11** (3):101–10.
38 Banning M. The think-aloud approach as an educational tool to develop and assess clinical reasoning in undergraduate students. *Nurse Educ Today* 2008;**28** (1):8–14.
39 Croskerry P. Clinical cognition and diagnostic error: applications of a dual process model of reasoning. *Adv Health Sci Educ Theory Pract* 2009;**14** (Suppl 1): 27–35.
40 Malterud K. Qualitative research: standards, challenges, and guidelines. *Lancet* 2001;**358** (9280):483–8.
41 Holden JD. Hawthorne effects and research into professional practice. *J Eval Clin Pract* 2001;**7** (1): 65–70.
42 Bryman A, Bell EA, Teevan JJ. *Social Research Methods*, 2nd edn. Toronto, ON: Oxford University Press 2009;564–89.
43 Silverman D. *Doing Qualitative Research: A Practical Handbook*, 3rd edn. London: Sage Publications 2010;121–2.
44 Bhutta ZA, Chen L, Cohen J *et al.* Education of health professionals for the 21st century: a global independent commission. *Lancet* 2010;**375** (9721):1137–8.
45 Bateman J, Allen ME, Kidd J, Parsons N, Davies D. Virtual patients design and its effect on clinical reasoning and student experience: a protocol for a randomised factorial multi-centre study. *BMC Med Educ* 2012;**12** (1):62.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Table S1. Design of cases.

Appendix S1. Virtual Patient XML case files used in the research.

Received 11 September 2012; editorial comments to authors 24 October 2012, accepted for publication 10 December 2012

Appendix 8 b.)

Bateman, J., Allen, M. E., Kidd, J., Parsons, N. & Davies, D. 2012c. Virtual patients design and its effect on clinical reasoning and student experience: A protocol for a randomised factorial multi-centre study. *BMC Medical Education,* 12, 62.

Bateman *et al. BMC Medical Education* 2012, **12**:62
http://www.biomedcentral.com/1472-6920/12/62

BMC Medical Education

STUDY PROTOCOL **Open Access**

Virtual patients design and its effect on clinical reasoning and student experience: a protocol for a randomised factorial multi-centre study

James Bateman[1,2*], Maggie E Allen[2], Jane Kidd[1], Nick Parsons[3] and David Davies[1]

Abstract

Background: Virtual Patients (VPs) are web-based representations of realistic clinical cases. They are proposed as being an optimal method for teaching clinical reasoning skills. International standards exist which define precisely what constitutes a VP. There are multiple design possibilities for VPs, however there is little formal evidence to support individual design features. The purpose of this trial is to explore the effect of two different potentially important design features on clinical reasoning skills and the student experience. These are the branching case pathways (present or absent) and structured clinical reasoning feedback (present or absent).

Methods/Design: This is a multi-centre randomised 2x2 factorial design study evaluating two independent variables of VP design, branching (present or absent), and structured clinical reasoning feedback (present or absent). The study will be carried out in medical student volunteers in one year group from three university medical schools in the United Kingdom, Warwick, Keele and Birmingham. There are four core musculoskeletal topics. Each case can be designed in four different ways, equating to 16 VPs required for the research. Students will be randomised to four groups, completing the four VP topics in the same order, but with each group exposed to a different VP design sequentially. All students will be exposed to the four designs. Primary outcomes are performance for each case design in a standardized fifteen item clinical reasoning assessment, integrated into each VP, which is identical for each topic. Additionally a 15-item self-reported evaluation is completed for each VP, based on a widely used EViP tool. Student patterns of use of the VPs will be recorded.
In one centre, formative clinical and examination performance will be recorded, along with a self reported pre and post-intervention reasoning score, the DTI. Our power calculations indicate a sample size of 112 is required for both primary outcomes.

Discussion: This trial will provide robust evidence to support the effectiveness of different designs of virtual patients, based on student performance and evaluation. The cases and all learning materials will be open access and available on a Creative Commons Attribution-Share-Alike license.

Keywords: Virtual patients, Clinical reasoning, Elearning, Education, Undergraduate, Musculoskeletal, Rheumatology

* Correspondence: james.bateman@warwick.ac.uk
[1]Education and Development Research Team, Warwick Medical School, Coventry CV4 7AL UK
[2]University Hospitals Coventry and Warwickshire, NHS Trust, Lakin Road, Coventry CV2 2DX UK
Full list of author information is available at the end of the article

Background

Virtual patients

Virtual patients(VPs) can be defined as electronic representations of realistic clinical cases [1]. They have been proposed as being an ideal tool to teach clinical reasoning skills [2]. A recent literature review, and systematic review of the literature has highlighted a lack of evidence supporting individual design properties for virtual patients as in other elearning areas [3]. VPs are widely used in up to one third of US and Canadian medical schools, however until 2007 development costs were high [4]. Multiple tools now exist to author virtual patients Much of the focus until now was been on their utility as educational tools in comparison to traditional teaching, in keeping with other elearning research [5].

A range of software packages exist for case authoring including 'CAMPUS', University of Heidelberg [6]; 'Labyrinth' from the University of Edinburgh [7]; 'Web-SP' from the Karolinska Institute, Sweden [8]; and 'vpSim' from the University of Pittsburgh [9]. Other researchers used bespoke software solutions to author cases [10,11].

An international technology interoperability standard for VPs was adopted in 2009 by the Medbiquitous Consortium [12]. This benchmark allows the interoperability of VP cases between compatible software systems allowing authoring, editing and playing of cases [13]. This facilitates collaboration, research, open access and the upkeep of these electronic resources [14,15]. This has potentially changed the working definition of what a VP is, by re-defining properties and dimensions of VPs [16,17]. A European Commission funded study has produced self-reported evaluation scores to help evaluation, the EViP project [15].

There are numerous VP design properties identified in the literature [2] and their impact on the learning experience are poorly understood[3]. Of particular interest are, firstly, the use of branching case pathways [18], and secondly, the role of structured feedback to promote clinical reasoning. Branching cases are more difficult to construct, more expensive when compared with linear cases [4], and may have unpredictable effects on individual students [19]. Clinical reasoning teaching support provides structured approaches to clinical reasoning [20,21], which can be deployed in VPs. In the 'SNAPPS' approach [20], students summarise case findings, narrow a differential diagnosis, analyse the differential diagnosis by comparing and contrasting possibilities, and plan management.

A number of validated tools exist to evaluate clinical reasoning and student experiences with a VP. To measure clinical reasoning, tools include the Key Feature Problem [22,23], and the Diagnostic Thinking Inventory [24,25]. Other appropriate assessments include multiple-choice questions, Bayesian reasoning questions [25,26] and diagnostic proficiency [20].

In musculoskeletal medicine there is a challenge in meeting the needs of increasing student populations [27]. This is confounded by a lack of exposure to clinical cases [28] which can potentially be mitigated by the use of virtual patients.

Problem statement and hypothesis

The influence of different design features in a VP is under-researched, although they can significantly affect the time and cost of VP production. A research study in this area would be able to answer this research question.

We hypothesise that the two important independent VP design variables, branching (present or absent), and structured clinical reasoning feedback (present or absent) are likely to influence students clinical reasoning in cases, and their user experience in terms of realism, engagement and learning value.

Methods

Objectives

The aim of this study is to evaluate how two independent VP design variables influence their effectiveness as an educational tool in musculoskeletal medicine. The specific objectives in the study are firstly to evaluate the performance of students exposed to different virtual patient designs in identical assessments of clinical reasoning skills. Secondly we aim to determine how different VP designs influence the student experience when using a VP. Finally we are attempting to explore the relationships between student performance in VP assessment metrics and other measurements of clinical skills, including written and clinical examinations.

Study design

This is a randomised 2x2 factorial design study evaluating two independent variables of VP design, branching (present or absent), and structured clinical reasoning feedback (present or absent).

Setting and participants

The setting is three university medical schools in the United Kingdom. These are the Warwick Medical School (WMS), the University of Birmingham Medical School (UBMS), and Keele Medical School (KMS). WMS runs a four year MBChB degree open only to graduate entry medical students, UBMS and KMS have a five year MBChB degree course, open to undergraduate entry medicine (UEM) graduate entry medical (GEM) students. The research project will run from 2011 to 2013.

Virtual Patient software information technology

Virtual patient cases in the study are created to the Medbiquitous standard [12] using the XML programming language [29]. The software used to create and host the cases is DecisionSim® v2.0, developed by the University of Pittsburgh. The cases are compatible with open source VP systems such as Open Labyrinth [7]. Access to cases, participation, electronic consent, and post case evaluations will be controlled by the VP software, and content hosted on the University of Warwick virtual learning environment Internet pages. Students will be registered with and logged in to the software, allowing tracking of decisions and performance.

Randomisation

The study follows the CONSORT statement on randomised trials [30]. A flow diagram of the study design is seen in Figure 1. Students from the eligible year-groups in each institution will be allocated to one of four intervention groups using block randomisation. Each of the university cohorts will be randomised individually. Block randomisation will use a computerised random number generator to allocate students. The primary investigator (JB) will implement the allocation and hold a record of the sequence.

Recruitment and baseline data collection

All eligible students will be invited to participate in the study. Inclusion criteria are students in the year group studying MSK medicine. The only exclusion criteria are students who do not volunteer or consent. Eligible students will be invited to attend an oral presentation and demonstration of the study, and given an approved study participant information sheet. Students who do not electronically record their informed consent will not be able to complete any cases, and are not considered to be participants. Students who consent will be considered to be study participants from this point onwards. At this point the baseline data collected from students will be gender, email address, student type (UEM/ GEM), year of study, and institution.

Additional data and information on other aspects of student performance will be collected from the examinations officer at WMS only. This includes student performance on formative clinical and written assessments at both at the end of the musculoskeletal block, and the end of year assessments.

Intervention and Independent design variables

The intervention consists of students completing four VP cases sequentially. Each case takes approximately 30 minutes to complete. The cases focus on four core clinical musculoskeletal areas. These are large joint arthritis, back pain, polyarthritis, and connective tissue disease. The 2x2 factorial study design means that any cases can be designed in four different ways. The four case designs are: A) not branched+no-feedback; B) branched+no feedback; C) not branched+feedback; D) branched +feedback. Students will use all four of the case designs during the research (see Figure 1).

The first variable is branching pathways through the VP, present or absent. There are four branching points with three choices through the thirty-minute case. This gives a possible 81 core pathways (3^4) through the case in a branched form. The linear case has a single core pathway, with participants being redirected back to the core pathway irrespective of the decision made, for example by feedback from a supervising clinician in the case. The second variable is the use of structured feedback to promote clinical reasoning skills, present or absent. This will be in a predetermined approach through the cases at five key points through the case, based on the 'SNAPPS' approach [20], systematic approaches to help Bayesian reasoning [21], and symptom categorisation [31].

Cases will be piloted and tested by healthcare professionals and a cohort of students in one centre prior to the study commencing. For the study, students will complete cases at WMS and KMS is in the form of sequential teaching sessions to students, taking place in a computer cluster. Students at UBMS will complete cases during allocated self-study time during their musculoskeletal block.

Other than the described independent variables we will control for other design variables highlighted in a critical literature review [2].

Inclusion and exclusion criteria

Inclusion criteria are students enrolled on the medical degree course and in the musculoskeletal teaching block in one of the medical schools in the study. Students must electronically sign consent to be included. Exclusion criteria are students from other year-groups. Students registered for a medical degree are required and assumed to have appropriate language and information technology skills.

Blinding

Students will be blind to their group allocation. Investigator blinding for the purposes of the data analysis and allocation is not used. In the institution where clinical examination performance is recorded, none of the investigators examine within the clinical specialty (musculoskeletal medicine).

Outcome measures

The primary outcome measures in this study are the performance in standardised composite clinical reasoning assessment using validated tools, and a modified self reported 15-item evaluation, reviewing four domains. These will be completed both during and

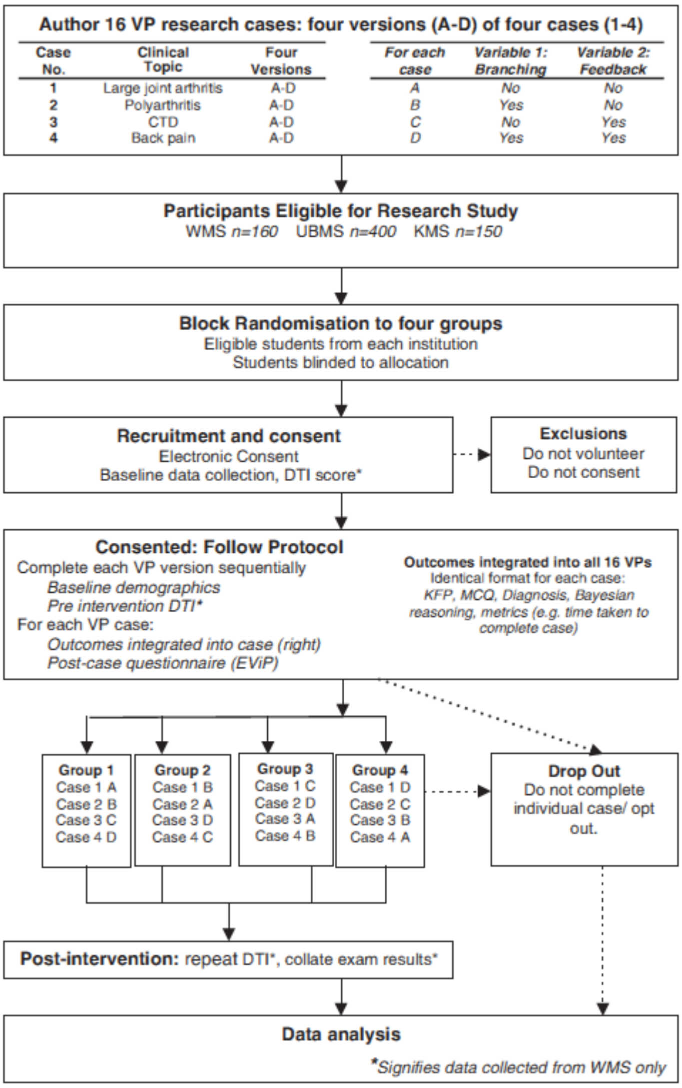


Figure 1 Flow diagram of study Proposed study protocol.

immediately following each case. The secondary outcome measures are engagement and patterns of use within the cases and collected from the online environment (see Table 1). For each case, the composite clinical reasoning assessment consists of validated assessment of clinical reasoning. These 15 items are eight 'key feature problem' questions, one Bayesian reasoning question, two multiple choice questions on diagnosis, and four multiple choice clinical decision questions. For each case the content of these 15 items is identical, allowing comparison between case designs. This allows comparison of a case which is not branched with no structured feedback vs. the same case in a branched format with structured feedback.

Student evaluation of each case will be collected using an electronic version of the EViP questionnaire, a fifteen item self reported evaluation. This explores exploring authenticity, professionalism, learning, and coaching through the case, using Likert scales with additional free text responses. Secondary outcome measures for each case are student's patterns of use of the case, such as time taken per case, case completion rates, and time taken to complete individual decisions.

Additional data will be collected from one centre, WMS, to support the validity of the VPs as

Table 1 Outcomes Measured during the study

Institution	*Outcomes for individual cases*	*Timing*
	All institions (WMS, KMS, UBMS)	
Primary Outcome Measures collected for each VP		
Validated clinical reasoning assessments.	Key Feature Problems score (x/8)	Student completes during the case
	Clinical Decision (x/4)	
	Bayesian Statistical Question (x/1)	
	Working diagnosis (x/2)	
	(Total score x/15)	
Self reported evaluation (EVIP)	*EViP Questionnaire (multiple domains)*	After each case
	Case Preference: reasoning (n from 4) learning (n from 4)	On completion of all cases
	Preference of case (learning)	
	Preference of case (realism)	
Secondary Outcome measures		
Other metrics collected in electronic environment	Time spent per case (seconds)	During the case, recorded automatically
	Time spent per node (seconds)	
	Number of nodes visited	
	Case completion percentage.	
	Time spent per question (seconds)	
WMS Only		
Validated self reported reasoning assessment (DTI)	Baseline and Mean improvement	Immediately Pre-and post- intervention
Summative assessment	Written and clinical	1 week post intervention
Formative assessment	Written and clinical	End of the year

educational and assessment tools. This includes a pre- and post-test Diagnostic Thinking Inventory, a 41 item validated assessment of clinical reasoning ability [24]. Performance in summative and formative written and clinical assessments, measured one week following the VP case, and several months later will also be collected.

Sample size determination

The authors agreed an important educational effect of a 5% difference in the score on validated assessments of clinical reasoning skills, and student self reported evaluations score. As no gold standard exists for the measurement of clinical reasoning skills, we have based the sample size calculation on performance for clinical reasoning on performance in the key feature problems (KFPs) integrated into each VP case. A previous study has shown mean KFP scores in a student population to be approximately normally distributed, with a standard deviation of 1.32 [32]. In this study where we will use 16 KFPs, a 5% difference in scores is considered significant, that is a difference in mean scores of 0.8, corresponding to a standardised effect size of approximately 0.6 (moderate to large). Based on these assumptions, we would require a total sample size of 88 students to detect this difference with 80% power at the (two-sided) 5% level. Assuming the effect size to be the same for both branching and feedback interventions, a sample size of 88 students would provide sufficient power to detect the main effects and an interaction effect that was twice as large as the assumed main intervention effect in the setting shown in Table 2.

If the interaction between branching and feedback interventions is of same order of magnitude as the expected main effects then we would require a fourfold increase in the sample size to give a total of 352 students. [33]

For self-reported scores, where a previous study reported a standard deviation of 0.93, [34] a 10% difference in scores (with a maximum of 5) is considered significant, that is a difference in mean scores of 0.5, corresponding to a standardised effect size of approximately 0.5 (moderate). Based on these assumptions, we would require a total sample size of 112 students to detect this difference with 80% power at the (two-sided) 5% level (Table 2).

Therefore 112 students would be required to detect the main effects and large interaction effect of branching and feedback on self-reported scores. To detect an interaction effect of the same order of magnitude as the expected main effects would require a total of 448 students. The pool of students available for recruitment

Table 2 Sample size calculation for Key Feature Problems outcomes, and student self evaluation scores

Key Feature problems					Student self reported Evaluation scores				
		Branching		*Total*			Branching		Total
		No	Yes				No	Yes	
Feedback	No	22	22	*44*	Feedback	No	28	28	*56*
	Yes	22	22	*44*		Yes	28	28	*56*
Total		*44*	*44*	***88***	*Total*		56	*56*	***112***

into this study is large at the three centres (WMS, n~160; UBMS, n~400; KMS, n~150). Given unforeseen recruitment problems and some loss to follow-up, a target of 112 students should be easily achievable to quantify the main intervention effects (branching and feedback) which are the primary focus of the study, with increasing recruitment above this target providing increasing power to detect potential interactions between the main effects.

Data analysis

We will present absolute numbers for enrolment, eligibility, and complete follow up. Descriptive statistics will be used to present student demographics, along with the mean, standard deviation, standard error of the mean, and 95% confidence intervals for primary and secondary outcome measures.

The primary analysis will be based on complete cases on a per-protocol analysis. It seems likely that some data may not be available due to voluntary withdrawal of participants, or drop-out through lack of completion of individual data items , unforeseen technical difficulties, and general loss to follow-up. Where possible the reasons for data 'missingness' will be ascertained and reported. The pattern of the missingness will be carefully considered and the reasons for non-compliance, withdrawal or other protocol violations will be stated and any patterns summarised. The primary analysis will investigate the fixed effects of the factorial combinations of branching and feedback on the primary outcome measures, performance in a standardised composite clinical reasoning assessment and a 15-item self reported evaluation. Analysis of covariance (ANCOVA) will be used to identify main effects, effect sizes, and interactions between the two independent design variables (feedback and branching). Blocking factors in the ANCOVA will adjust for the effects randomisation group, case ordering and recruiting centre, with student GEM status and gender as covariates. Tests from the ANCOVA will be two-sided and considered to provide evidence for a significant difference if p-values are less than 0.05 (5% significance level). Estimates of treatment effects will be presented with 95% confidence intervals. Students case preferences for learning and realism, and EViP will be evaluated using chi-squared tests for grouping factors case design and number.

We will determine the predictive validity of performance in the VP composite assessment, using one institution's summative examination results, WMS. We will use the correlation coefficient (Pearson's product–moment, r) to determine the effect size of any linear correlation between the VP scores and institution examinations.

A detailed statistical analysis plan (SAP) will be agreed with the trial management group at the start of the study, with any subsequent amendments to this initial SAP being clearly stated and justified. The routine statistical analysis will mainly be carried out using R (http://www.r-project.org/) and S-PLUS (http://www.insightful.com/). Results from this study will also be compared with results from other studies.

Ethical approval

The National Health Service Local Research Ethics Committee approved the protocol as an educational research study. Warwick Medical School Biomedical Research Ethics Committee gave written ethics and institution approval in 2010. The study has institutional approval from KMS and UBMS.

Discussion

The main purpose of this randomised-factorial study design is to identify the most effective design principles for VPs across a range of musculoskeletal cases, addressing a research question identified recently in the literature [2], which have not been answered by a recent meta-analysis [3].

Our use of validated assessments of clinical reasoning where possible helps to validate the research findings, however there are limitations in these existing tools to measure clinical reasoning. The use of assessment data from one institution from both summative and written examinations will assist with the interpretation of the predictive and criterion validity of the VP cases when considered as formative assessments.

The blinding of students to group allocations will hopefully minimise bias and preconceptions about virtual patients. The students in the study do not have VP education formally integrated into any curriculum, however previous exposure and familiarity with existing open access cases cannot be excluded.

Interpretation of the research findings will be facilitated by open access publication of VPs generated. This has not been used in recent published peer reviewed research on VPs [2,3,34-36]. The publication of these cases with the research will allow appropriate integration of the materials as a learning resource of the design process, and allow other researchers to evaluate the research methods [37].

Competing interests
The authors' declare that they have no competing interests.

Authors' contributions
All of the authors have had a substantial contribution to the research design and study protocol. JB and DD conceived the original study design, which was reviewed by MA and JK. NP advised on the statistical analysis and power calculations. JB wrote the initial draft, which was revised and approved by MA, JK, NP and DD for content. All authors have approved the final draft of this work for publication.

Authors' information
JB is an Education Research Fellow with Arthritis Research UK, undertaking a PhD at Warwick Medical School. MA is a consultant rheumatologist and Director of Medical Education at University Hospitals Coventry and Warwickshire NHS Trust, where NP works as a statistician with Warwick Medical School. JK is head of the Education Development and Research Team (EDRT), Warwick Medical School. DD is Associate Professor of Medical Education at the EDRT.

Acknowledgments
This study is funded from an Educational Research Fellowship Grant from Arthritis Research UK, UK Registered charity number 207711. We would like to thank Arthritis Research UK Education and Training Committee, and faculty from musculoskeletal departments at the host and collaborating institutions. We wish to also thank the publisher Wiley for the use of the Diagnostic Thinking Inventory. We thank Dr JB McGee from the University of Pittsburgh, and DecisionSim for assistance and collaboration with the VP production.

Author details
[1]Education and Development Research Team, Warwick Medical School, Coventry CV4 7AL UK. [2]University Hospitals Coventry and Warwickshire, NHS Trust, Lakin Road, Coventry CV2 2DX UK. [3]Department of Statistics, Warwick Medical School, Coventry CV4 7AL UK.

Received: 28 May 2012 Accepted: 4 July 2012
Published: 1 August 2012

References
1. Ellaway RH, Poulton T, Smothers V, Greene P: **Virtual patients come of age.** *Med Teach* 2009, **31**(8):683–684.
2. Cook DA, Triola MM: **Virtual patients: a critical literature review and proposed next steps.** *Medical Education* 2009, **43**(4):303–311.
3. Cook DA, Erwin PJ, Triola MM: **Computerized virtual patients in health professions education: a systematic review and meta-analysis.** *Acad Med* 2010, **85**(10):1589–1602.
4. Huang G, Reynolds R, Candler C: **Virtual patient simulation at US and Canadian medical schools.** *Acad Med* 2007, **82**(5):446–451.
5. Cook DA, Levinson AJ, Garside S, Dupras DM, Erwin PJ, Montori VM: **Internet-based learning in the health professions: a meta-analysis.** *Jama* 2008, **300**(10):1181–1196.
6. Ruderich F, Bauch M, Haag M, Heid J, Leven FJ, Singer R, Geiss HK, Junger J, Tonshoff B: **CAMPUS–a flexible, interactive system for web-based, problem-based learning in health care.** *Studies in health technology and informatics* 2004, **107**(Pt 2):921–925.
7. Begg M: **Virtual patients: practical advice for clinical authors using Labyrinth.** *Clin Teach* 2010, **7**(3):202–205.
8. Zary N, Johnson G, Boberg J, Fors UG: **Development, implementation and pilot evaluation of a Web-based Virtual Patient Case Simulation environment–Web-SP.** *BMC medical education* 2006, **6**:10.
9. McGee JB, Neill J, Goldman L, Casey E: **Using multimedia virtual patients to enhance the clinical curriculum for medical students.** *Studies in health technology and informatics* 1998, **52**(Pt 2):732–735.
10. Wilson AS, Goodall JE, Ambrosini G, Carruthers DM, Chan H, Ong SG, Gordon C, Young SP: **Development of an interactive learning tool for teaching rheumatology–a simulated clinical case studies program.** *Rheumatology (Oxford)* 2006, **45**(9):1158–1161.
11. Friedman CP, France CL, Drossman DD: **A randomized comparison of alternative formats for clinical simulations.** *Med Decis Making* 1991, **11**(4):265–272.
12. *MedBiquitous Virtual Patient Player Specifications and Description Document V.1.0*: http://www.medbiq.org/working_groups/virtual_patient/VirtualPatientPlayerSpecification.pdf.
13. Ellaway R, Poulton T, Fors U, McGee JB, Albright S: **Building a virtual patient commons.** *Med Teach* 2008, **30**(2):170–174.
14. Ellaway RH, Davies D: **Design for learning: deconstructing virtual patient activities.** *Med Teach* 2011, **33**(4):303–310.
15. Smothers V, Ellaway R, Balasubramaniam C: **eViP: sharing virtual patients across Europe.** *AMIA Annual Symposium proceedings / AMIA Symposium AMIA Symposium* 2008, **1140**.
16. Bateman J, Davies D: **Virtual patients: are we in a new era?** *Acad Med* 2011, **86**(2):151. author reply 151.
17. Triola MM, Campion N, McGee JB, Albright S, Greene P, Smothers V, Ellaway R: **An XML standard for virtual patients: exchanging case-based simulations in medical education.** *AMIA Annual Symposium proceedings / AMIA Symposium AMIA Symposium* 2007, **2007**:741–745.
18. Huwendiek S, De leng BA, Zary N, Fischer MR, Ruiz JG, Ellaway R: **Towards a typology of virtual patients.** *Med Teach* 2009, **31**(8):743–748.
19. Bateman J, Davies D, Allen M: **Mind wandering has an impact in electronic teaching cases.** *Medical Education* 2012, **46**(2):235.
20. Wolpaw T, Papp KK, Bordage G: **Using SNAPPS to facilitate the expression of clinical reasoning and uncertainties: a randomized comparison group trial.** *Acad Med* 2009, **84**(4):517–524.
21. Sedlmeier P, Gigerenzer G: **Teaching Bayesian reasoning in less than two hours.** *J Exp Psychol Gen* 2001, **130**(3):380–400.
22. Page G, Bordage G, Allen T: **Developing key-feature problems and examinations to assess clinical decision-making skills.** *Acad Med* 1995, **70**(3):194–201.
23. Bordage G, Brailovsky C, Carretier H, Page G: **Content validation of key features on a national examination of clinical decision-making skills.** *Acad Med* 1995, **70**(4):276–281.
24. Bordage G, Grant J, Marsden P: **Quantitative assessment of diagnostic ability.** *Medical Education* 1990, **24**(5):413–425.
25. Round AP: **Teaching clinical reasoning–a preliminary controlled study.** *Medical Education* 1999, **33**(7):480–483.
26. Sumner W 2ND, Hagen MD: **Value of information in virtual patient performance evaluations.** *AMIA Annual Symposium proceedings / AMIA Symposium AMIA Symposium* 2008, **1149**.
27. Badcock LJ, Raj N, Gadsby K, Deighton CM: **Meeting the needs of increasing numbers of medical students–a best practise approach.** *Rheumatology (Oxford)* 2006, **45**(7):799–803.
28. Adebajo A, Windsor K, Hassell A, Dacre J: **Undergraduate education in rheumatology.** *Rheumatology (Oxford)* 2005, **44**(9):1202–1203. author reply 1203.
29. *Extensible Markup Language (XML) 1.0*. http://www.w3.org/TR/REC-xml/-sec-origin-goals.
30. Moher D, Schulz KF, Altman DG: **The CONSORT statement: revised recommendations for improving the quality of reports of parallel group randomized trials.** *BMC Med Res Methodol* 2001, **1**:2.
31. Norman GR, Eva KW: **Diagnostic error and clinical reasoning.** *Medical Education* 2010, **44**(1):94–100.
32. Fischer MR, Kopp V, Holzer M, Ruderich F, Junger J: **A modified electronic key feature examination for undergraduate medical students: validation threats and opportunities.** *Med Teach* 2005, **27**(5):450–455.
33. Montgomery AA, Peters TJ, Little P: **Design, analysis and presentation of factorial randomised controlled trials.** *BMC Med Res Methodol* 2003, **3**:26.

34. Berman N, Fall LH, Smith S, Levine DA, Maloney CG, Potts M, Siegel B, Foster-Johnson L: **Integration strategies for using virtual patients in clinical clerkships.** *Acad Med* 2009, **84**(7):942–949.
35. Botezatu M, Hult H, Fors UG: **Virtual patient simulation: what do students make of it? A focus group study.** *BMC medical education* 2010, **10**:91.
36. Zary N, Johnson G, Fors U: **Web-based virtual patients in dentistry: factors influencing the use of cases in the Web-SP system.** *Eur J Dent Educ* 2009, **13**(1):2–9.
37. Wilbanks J: **Another reason for opening access to research.** *Bmj* 2006, **333** (7582):1306–1308.

doi:10.1186/1472-6920-12-62
Cite this article as: Bateman *et al.*: **Virtual patients design and its effect on clinical reasoning and student experience: a protocol for a randomised factorial multi-centre study.** *BMC Medical Education* 2012 **12**:62.

Appendix 8 c.)

Bateman, J. & Davies, D. 2011. Virtual patients: Are we in a new era? *Academic Medicine : Journal of the Association of American Medical Colleges,* 86, 151; author reply 151.

Letters to the Editor

Virtual Patients: Are We in a New Era?

To the Editor: We read with interest the comprehensive systematic review by Cook et al[1] providing the first substantive quantitative review evidence supporting the educational impact of virtual patients (VPs). While this review summarizes well the historical uses of VPs, we would like to highlight important changes to VP design and authorship that have implications for the interpretation of the review article.

The review includes computerized patient cases from 1991. Although cases both pre- and post-1991 in the review may be by definition *virtual patients*, we argue that this arbitrary cutoff introduces a historical bias obscuring a more recent VP technology shift. Advances in software during the past five years have brought about a change of direction in VP development toward clinicians being the primary authors, designers, editors, publishers, and distributors of cases.[2] The MedBiquitous Virtual Patient interoperability standard introduced in 2009 has facilitated the sharing of VP cases and collaboration, resulting in a standardized point of reference for VP design and perhaps a tighter definition of what a VP is.[3] Consequently, all major VP authoring systems now use exclusively Web-based authoring and editing. With clinicians now able to directly edit, restructure, and repurpose cases, the authenticity of cases may be enhanced, with the opportunity to adapt branching structures as needed. In practical terms these changes have led to international collaboration in open-access research projects (such as the European Union's cofunded Electronic Virtual Patient Programme, eViP), lower development costs, and a shift toward clinician-centered VPs in our own student curricula.[4]

In terms of Cook and colleagues' review, we therefore suggest that the current generation of VPs being developed represents a significant shift from previous cases, and this should be carefully considered when interpreting evidence from the older, non-Web-based commissioned computerized clinical cases. We propose that against the backdrop of these new technology advances, further weight should be added to the call from Cook et al for robust research into theory-based comparisons of current VPs.

James Bateman, MBChB

Education research fellow, Institute of Clinical Education, Warwick Medical School, Coventry, United Kingdom; james.bateman@warwick.ac.uk.

David Davies, PhD

Associate professor in medical education and elearning, Institute of Clinical Education, Warwick Medical School, Coventry, United Kingdom.

References

1 Cook DA, Erwin PJ, Triola MM. Computerized virtual patients in health professions education: A systematic review and meta-analysis. Acad Med. 2010;85:1589–1602.
2 Huwendiek S, De leng BA, Zary N, Fischer MR, Ruiz JG, Ellaway R. Towards a typology of virtual patients. Med Teach. 2009;31:743–748.
3 MedBiquitous Web site. http://www.medbiq.org/std_specs/standards/index.html. Accessed October 5, 2010.
4 eViP: Electronic Virtual Patients Web site. http://www.virtualpatients.eu. Accessed October 3, 2010.

In Reply: We agree with Dr. Bateman and Dr. Davies that there have indeed been significant advances over the past 45 years in the technology used to implement virtual patients (VPs). However, the fundamental premise has not changed. Our review[1] found that many early systems were amazingly prescient about the functionality, design, and pedagogy of modern VP applications, including those using the recent MedBiquitous VP (MVP) standard.[2] For example, we found sophisticated branching VPs dating back as far as 1973,[3] and most of the VPs we identified were actually authored by clinicians.

We maintain, however, that it is the instructional and narrative design—not the hardware or software—that fundamentally influences learning. Unfortunately, in these critical qualities VPs have not advanced much in 40 years.

We also agree that the environment for meaningful theory-based VP research is better than ever. As we discussed in a previous article,[4] standards such as MVP, repositories of shared cases, and the availability of free VP players and authoring systems such as OpenLabyrinth[5] will facilitate the use of these tools across the spectrum of medical education, while collaborations such as the European eViP project[6] offer new models for distributing and evaluating VPs. Such platforms have the potential to become multiinstitutional research laboratories, with the learner volume and diversity needed to rigorously answer key questions about the design and use of VPs. These innovations reflect, in one sense, a more fundamental advance than the previously incremental improvements in technology over the past several decades. We hope and trust that these new advantages will pave the way to more and better research in the field.

Marc M. Triola, MD

Assistant professor of medicine and director, Division of Educational Informatics, New York University School of Medicine, New York, New York; marc.triola@nyumc.org.

David A. Cook, MD, MHPE

Professor of medicine and medical education and director, Office of Education Research, Mayo Clinic College of Medicine, Rochester, Minnesota.

References

1 Cook DA, Erwin PJ, Triola MM. Computerized virtual patients in health professions education: A systematic review and meta-analysis. Acad Med. 2010;85:1589–1602. http://journals.lww.com/academicmedicine/Abstract/2010/10000/Computerized_Virtual_Patients_in_Health.16.aspx. Accessed October 20, 2010.
2 MedBiquitous Virtual Patient Specifications and Description Document. http://www.medbiq.org/working_groups/virtual_patient/VirtualPatientDataSpecification.pdf. Accessed October 20, 2010.
3 Hoffer EP. Experience with the use of computer simulation models in medical education. Comput Biol Med. 1973;3:269–279.
4 Cook DA, Triola MM. Virtual patients: A critical literature review and proposed next steps. Med Educ. 2009;43:303–311.
5 OpenLabyrinth. http://sourceforge.net/projects/openlabyrinth. Accessed October 20, 2010.
6 eViP Web site. http://www.virtualpatients.eu. Accessed October 20, 2010.

Appendix 8 d.)

Bateman, J., Davies, D. & Allen, M. 2012d. Mind wandering has an impact in electronic teaching cases. *Medical Education,* 46, 235.

Mind wandering has an impact in electronic teaching cases

James Bateman,[1,2] David Davies[1] & Maggie Allen[2]

Editor – We read with interest the article by Smallwood *et al.*[1] and the accompanying commentaries by Hilton[2] and Berner[3] regarding mind wandering in clinical medicine. We believe the principles proposed are equally applicable to e-learning instruction that uses clinical cases in the form of 'virtual patients' (VPs) and have evidence to support the impact of mind wandering on learning and simulated student performance.

Virtual patients can be defined as the 'interactive computer simulation of clinical scenarios for the purpose of medical training and assessment',[4] and are perhaps best placed to teach clinical reasoning skills.[5] The evidence to support individual design features is limited.[6] In an ongoing mixed-methods research project into design principles for VP cases, we have identified phenomena that parallel mind wandering in real clinical performance. We have conducted a series of six focus groups using grounded theory methodology with students who previously completed two VP cases, each of which was 30 minutes long. Mind wandering recurred as a theme across the focus groups in different gender, age and student year groups. Our preliminary theory suggests that mind wandering when using VPs results from a complex interplay among the underlying conditions or principles by which VPs are constructed, contextual factors and the net impact these have on the student–VP interaction. The three main conditional factors are the underlying clinical content, e-learning object design, and pedagogic approaches. These result in a series of student–VP interactions and appear to influence relationships among students and the electronic characters, mental case building, task avoidance and subsequent mind wandering. This theory is supported by student 'in case' patterns of use and performance. Student performance measures include probabilistic decision making, diagnostic proficiency, and assessments using a validated method of assessing clinical reasoning skills.[7] To investigate these preliminary findings, we are undertaking a multicentre, randomised, factorial study looking at case design variables in a large cohort of medical undergraduates from three medical schools in the UK. This will look in particular at the influences of branching case pathways and the deliberate promotion of clinical reasoning skills through the cases.

In conclusion, we welcome the highlighting of the importance of mind wandering in the context of student case-based learning and the use of electronic VPs, and have evidence to support an impact on students in case performance. Ongoing research in the clinical and e-learning environments is likely to illuminate and potentially mitigate the problems posed by mind wandering.

[1]Institute of Clinical Education, Warwick Medical School, Warwick, UK
[2]Department of Medical Education, University Hospitals Coventry and Warwickshire NHS Trust, Coventry, UK

Correspondence: Dr James Bateman, Institute of Clinical Education, Warwick Medical School, Warwick, CV4 7AL, UK.
Tel: 00 44 247 574880; Fax: 00 44 247 6968791;
E-mail: james.bateman@warwick.ac.uk

doi: 10.1111/j.1365-2923.2011.04187.x

REFERENCES

1 Smallwood J, Mrazek MD, Schooler JW. Medicine for the wandering mind: mind wandering in medical practice. *Med Educ* 2011;**45**:1072–80.
2 Hilton S. Mind wandering in medical practice. *Med Educ* 2011;**45**:1066–7.
3 Berner ES. Mind wandering and medical errors. *Med Educ* 2011;**45**:1068–9.
4 Ellaway R, Poulton T, Fors U, McGee JB, Albright S. Building a virtual patient commons. *Med Teach* 2008;**30** (2):170–4.
5 Cook DA, Erwin PJ, Triola MM. Computerised virtual patients in health professions education: a systematic review and meta-analysis. *Acad Med* 2010;**85** (10):1589–602.
6 Cook DA, Triola MM. Virtual patients: a critical literature review and proposed next steps. *Med Educ* 2009;**43** (4):303–11.
7 Page G, Bordage G, Allen T. Developing key-feature problems and examinations to assess clinical decision-making skills. *Acad Med* 1995;**70** (3):194–201.